Exploring the regulation and activation of ALT,

the Alternative Lengthening of Telomeres

Thesis submitted for the degree of
Doctor of Philosophy
At the University of Leicester

By

Ahmed S. N. Alhendi

Department of Genetics
University of Leicester

2018

CONTENTS

Exploring the regulation and activation of ALT, the Alternative Lengthening of Telomeres

Ahmed S. N. Alhendi

Abstract

Cancers and cell lines that utilise the Alternative Lengthening of Telomeres (ALT) to maintain telomere length usually display complex karyotype rearrangements consistent with genome instability that may arise at crisis through dysfunctional telomeres. The absence of the ATRX chromatin remodelling factor is the predominant marker in these cancers and cell lines. It has been shown that specific orphan nuclear receptors (ORs) can bind to $(TCAGGG)_n$ repeats in telomeres of some ALT+ cell lines and this binding results in changes to the telomere chromatin that may facilitate the ALT.

In this project, the binding site and regulatory roles of ORs were investigated in a panel of five ALT+ cell lines. No relationship was identified between the expression of NR2F2 or NR2C2 proteins and the localisation of these proteins at telomeres in ALT+ cell lines. The frequency of cells with NR2F2 or NR2C2 foci at telomeres varies considerably between the ALT+ cell lines (7- 91% for NR2F2; 1-42% for NR2C2) with only the WI38VA13_2RA ALT+ cell line showing a high frequency for both, consistent with previous work. PCR amplification of XpYp, 12q and 17p telomeres followed by hybridisation to $(TTAGGG)_n$, $(TCAGGG)_n$ or $(TGAGGG)_n$ variant repeat probes was used to measure the frequency of telomere molecules that contain binding sites for these ORs. This revealed that some telomere alleles in ALT+ cell lines lack binding sites for these ORs, indicating that binding of these ORs is not essential for ALT at some telomeres. Downregulation of *NR2F2* in three ALT+ cells resulted in significant changes in the expression of genes involved in DNA replication and repair that resulted in cell cycle arrest in the U2OS, but not in the WV or WI38VA13_2RA ALT+ cell lines.

Integrated analyses of whole exome sequence (WES) and RNA-seq analyses were used to uncover the mutation and the gene expression changes associated with ALT activation after *ATRX* knockout, using CRISPR-Cas9, in SV40- transformed pre-crisis cells. Differential gene expression analysis at early post-crisis time-points in emerging clonal populations showed that downregulation of the JAK-STAT signalling pathway, the ALT-suppressor *SP100* and reduced expression of non-homologous end joining (NHEJ) related genes *NR4A1* and *XLF* were the most remarkable events. These gene expression changes were combined with copy number changes in chromosomes 8, 11 and 18 and remarkable level of genome instability. This could suggest that, following the loss of *ATRX* in SV40-transformed cells, the changes in the expression of the genes required for double strand break (DSB) repair contributed to the activation of ALT.

Acknowledgements

List of Abbreviations

APBs	ALT-associated PML nuclear bodies
ADD	ATRX-DNMT3-DNMT3L
a-EJ	Alternative end joining
AF1	Activation Function-1
AF2	Activation Function-2
ALT	Alternative Lengthening of Telomeres
ATM	Ataxia Telangiectasia Mutated
ATR	ATM- and Rad3-related
ATRX	Alpha-Thalassemia/mental Retardation X-linked
BIR	break-induced replication
BLM	Bloom syndrome RecQ like helicase
Chk2	cell cycle checkpoint 2
c-NHEJ	classical non-homologous end joining
CNV	Copy Number Variant
Cas9	CRISPR-Associated Sequence 9
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4',6-Diamidino-2-Phenylindole, Dihydrochloride
DAXX	the death domain associated protein
DBD	DNA-binding domain
DDR	DNA damage response
DE	differential expression
DMEM	Dulbeco's Modified Eagle Medium
DNA2	DNA Replication Helicase/Nuclease 2
DRs	direct repeat elements
DSB	double strand break
DSBR	Double strand break repair
ECTR	extrachromosomal telomeric repeat
FA	Fanconi anemia
FBS	Fetal Bovine Serum
FISH	fluorescence in situ hybridization
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
G4	G-quadruplex
DAVID	The Database for Annotation, Visualization and Integrated Discovery
GAPDH	glyceraldehyde-3 phosphate dehydrogenase
GO	Gene ontology
H3K9me3	Trimethylation of histone H3 at lysine 9
H4K20me3	Trimethylation of histone H4 at lysine 20
HP1	Heterochromatin protein 1
HR	homologous recombination
HREs	Hormone Response Elements
IF	ImmunoFluorescence
KEGG	Kyoto Encyclopaedia of Genes and Genomes
TERT	Human telomerase reverse transcriptase
LBD	Ligand-Binding Domain

LOH	Loss Of Heterozygosity
M1	Mortality barrier 1
M2	Mortality barrier 1
MDC1	Mediator of DNA-damage checkpoint 1
MEM	Minimum Essential Medium
MMEJ	microhomology-mediated end joining pathway
MRN	Mre11/Rad50/ Nbs1 complex
MS32	minisatellite locus 32 at chromosome 1q
NMD	nonsense-mediated decay
NR2C/F	Nuclear Receptor Subfamily 2, Groups C and F
NR2C1	Nuclear Receptor Subfamily 2, Group C, Member 1
NR2C2	Nuclear Receptor Subfamily 2, Group C, Member 2
NR2F1	Nuclear Receptor Subfamily 2, Group F, Member 1
NR2F2	Nuclear Receptor Subfamily 2, Group F, Member 2
NR2F6	Nuclear Receptor Subfamily 2, Group F, Member 6
NuRD	Nucleosome Remodelling and Histone Deacetylation factors
PDs	population doublings
ORs	Orphan nuclear receptors
PAM	protospacer adjacent motif
PCR	Polymerase chain reaction
PE	Paired end
PML	Promyelocytic Leukaemia nuclear bodies
POT1	Protection of Telomeres 1
RAP1	Repressor/activator protein 1
RNAi	The RNA interference
RPA	Replication Protein-A
RPMI	Roswell Park Memorial Institute Medium
RTEL1	Regulator of Telomere Elongation Helicase 1
SDSA	synthesis-dependent strand annealing
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SSA	Single strand annealing
STELA	Single Telomere Elongation Length Analysis
SV40-LT	large T antigen of Simian virus 40
TERRA	TElomeric Repeat-containing RNA
TIFs	telomere dysfunction induced foci
TIN2	TRF1-interacting nuclear protein 2
TMM	Telomere Maintenance Mechanism
TPP1	Tripeptidyl Peptidase I
TRF1	Telomeric Repeat binding Factor 1
TRF2	Telomeric Repeat binding Factor 2
T-SCE	telomere sister chromatid exchange
TTI	Targeted Telomere Insertions
TVR-PCR	Telomere variant repeat mapping by PCR

Chapter 1: Introduction

1.1 Telomeres

Telomeres are nucleoprotein structures located at the ends of linear chromosomes. The different properties of telomeres compared to other chromosomal regions were first studied in *Drosophila melanogaster* by Muller and in *Zea mays* by McClintock (Muller, 1938; McClintock, 1941). Muller observed unusual chromosome rearrangements in *Drosophila* cells following exposure to X-rays, from which he proposed the existence of specialised structures (Muller, 1938). In studying the stability of chromosomes in *Zea* cells, McClintock showed that the broken chromosomes in endosperm tissue are unstable and enter into breakage-fusion-bridge cycles in successive cell divisions. Moreover, McClintock noticed that the breakage-fusion-bridge cycles were associated with dicentric chromatid formation and aberrant recombination events. The breakage-fusion-bridge cycle ended after the zygote formation, where the broken chromosome was healed by the acquisition of a telomere (McClintock, 1941). These observations provided the initial evidence that the ends of chromosomes are important structures for maintaining genome stability by protecting the ends of chromosomes from degradation and unwanted recombination.

1.2 The structure of telomeres

The sequence of telomeres remained unknown until 1978 when the first telomeric sequence was identified in a unicellular ciliated protozoan *Tetrahymena thermophile* by Blackburn and Gall. Their work showed that telomeres in *T. thermophile* comprised a tandemly repeated sequence made of $(TTGGGG)_n$ (Blackburn and Gall, 1978). In ciliated protozoa, the G-reach strand is oriented 5' $\rightarrow$ 3' towards the end of the chromosome with a 3' single stranded overhang containing 12 to 16 nucleotides (nt) (Klobutcher *et al.*, 1981). Since then several studies have shown that telomere sequences are highly conserved across many species in the kingdom of eukaryotes (Meyne *et al.*, 1989). However, normal telomere length varies between organisms, for example 200-400 bp in

single cell organisms such as in *T. thermophile* (Blackburn and Gall, 1978) and *Saccharomyces cerevisiae* (Shampay *et al.*, 1984)*,* but much longer in mammals; 2-20 kb in *Homo sapiens* (human) (Moyzis *et al.*, 1988; Counter *et al.*, 1992), and up to 150 kb in *Mus musculus* (mice) (Starling *et al.*, 1990).

Human telomeres consist of a $(TTAGGG)_n$ double-stranded tandemly repeated sequence with the G-reach strand oriented 5' →3' towards of the end of the chromosome and forming a 3' single-strand overhang of 50 to 500 nt (Figure 1.1). The complementary C-reach strand predominantly ends with the sequence ATC-5'. This led to the proposal that after replication, a specific nuclease acts on the C-strand (Sfeir *et al.*, 2005). In addition to the canonical repeat, the proximal regions (0.5 -2 kb) of telomeres can include degenerate repeats such as TCAGGG, TTGGGG, and TGAGGG (Baird *et al.*, 1995; Varley *et al.*, 2002; Palm and de Lange, 2008).

1.2.1 Secondary structure of telomeres

1.2.2 T-loop formation

The telomeric 3' overhang can fold back and invade into double-stranded $(TTAGGG)_n$ DNA and base pair with the C-reach strand resulting in a looped structure called a telomere loop (t-loop). At the point of strand invasion, the telomeric G-strand is displaced forming a displacement loop (D-loop) (Figure 1.1 B). The t-loop was first visualised in mouse and human telomeres by electron microscopy (Griffith *et al.*, 1999). The closed structure of t-loops provides a protective cap for chromosome ends and hides the telomere from the DNA damage response (DDR) machinery (Griffith *et al.*, 1999).

Figure 1.1 Human Telomere Structure. (A) A schematic of degenerate variant repeats at the start of the telomere array, blue line represent G-strand, red line represent C-strand (B) t-loop and D-loop formation. (C) A model of the interactions between the telomere binding proteins. (D) A model of shelterin binding to a t-loop.

1.2.3 Chromatin structure of telomeres

Specific proteins bind to the TTAGGG repeats and regulate telomere length and function. The proteins can be classified into single-strand or double-strand binding proteins.

1.2.3.1 The shelterin complex

Human telomere binding proteins form a complex called shelterin. The shelterin complex consists of six proteins. The Telomeric Repeat binding Factor 1 (TRF1) and Telomeric Repeat binding Factor 2 (TRF2) bind to double stranded (TTAGGG)$_n$ repeats as homodimers. The telomeric repeat-binding factor 2-interacting protein 1 (RAP1) has no direct binding site to telomeric DNA, instead it forms a 1:1 complex with TRF2 (Li *et al.*, 2000; Celli and de Lange, 2005). The Protection of Telomeres 1 (POT1) protein binds to single stranded (TTAGGG)$_n$ (Palm and de Lange, 2008). The TRF1-interacting nuclear protein 2 (TIN2) has a central position in the shelterin complex, as it binds TRF1, TRF2, and its interacting subunit Tripeptidyl Peptidase I (TPP1). Consequently, TIN2 plays an important in maintaining the structure of shelterin complex, it also recruits TPP1 and therefore POT1 to telomeres (Palm and de Lange, 2008) (Figure 1.1 C).

As stated TRF2 binds to duplex (TTAGGG)$_n$ repeats where it inhibits the DNA damage response by Ataxia Telangiectasia Mutated (ATM) kinase signalling (Celli and de Lange, 2005). The depletion of TRF2 in mammalian cells results in the ends of telomeres being recognised as double strand breaks (DSB) that accumulate DNA damage response proteins such as γ-H2AX, p53 Binding Protein 1(53BP1) and the mediator of DNA-damage checkpoint 1 (MDC1) so forming telomere dysfunction induced foci (TIFs) (Takai *et al.*, 2003). This results in activation of ATM signalling, cell cycle checkpoint 2 (Chk2) and activation of p53. This DNA damage signalling normally results in p53/Rb-dependent senescence (Karlseder *et al.*, 1999) or ATM/p53-dependent apoptosis (Smogorzewska and de Lange, 2002). However, in the absence of p53 it results in NHEJ-mediated repair resulting in fused chromosomes that break in the following mitosis causing genome instability and eventually resulting in cell death (Celli and de Lange, 2005).

POT1 was first identified in human cells due to its sequence similarity to the single-stranded telomere binding TEBPα (the alpha subunit of the TEBPα/β) in *Oxytreacha nova.* Similar to TEBPα, the N terminus of POT1 consists of two Oligosaccharide-Binding (OB) domains, which allow it to bind to the single-stranded G-reach overhang (Baumann and Cech, 2001). When POT1 is recruit to telomeres, it forms a 1:1 heterodimer with TPP1. However, this TPP1-POT1 heterodimer seems to play diverse roles in the regulation of telomere elongation by telomerase. TPP1 can directly interact with telomerase and stimulate its activity, while the overloading of POT1 on G-reach overhangs was shown to limit access by telomerase (Xin *et al.*, 2007). Binding of POT1 to the G-reach overhangs also prevents a single-stranded DNA damage response by ATM- and Rad3-related (ATR) kinase signalling. In mouse embryonic fibroblasts, it has been shown that depletion of either of the Pot1α/β subunits was sufficient to activate the ATR kinase, Chk1/ Chk2, as well as p53. The depletion of POT1 or its partner TPP1 in p53$^{-/-}$ cells resulted in non-homologues-end-joining (NHEJ) mediated telomere-telomere fusion that lead to chromosomal instability (Denchi and de Lange, 2007). Thus, the shelterin complex is essential for telomere capping as it prevents the telomeric DNA from triggering DDR and DSBT machineries, thereby avoiding telomere fusions by NHEJ and unwanted recombination by homologous recombination (HR) (Figure 1.1 D).

1.2.3.2 The CST complex

The CST complex is a single-strand telomere binding complex which consists of three proteins, first identified in budding yeast (Cdc13, Stn1, Ten1) and then in mammalian cells (CTC1, STN1, and TEN1) (Nugent *et al.*, 1996; Hughes *et al.*, 2000; Miyake *et al.*, 2009). None of the proteins in the CST complex shows sequence similarity to POT1 or any of the other shelterin components, which suggests that the CST complex independently (Linger and Price, 2009). The loading of the CST complex onto telomeres is by the interaction of CTC1 with the G-reach single-stranded overhang through a single OB domain. Neither Stn1 nor Tent1 can directly bind to telomeres (Figure 1.1C) (Nugent *et al.*, 1996; Miyake *et al.*, 2009).

In cells that use telomerase, the CST complex regulates telomere lengthening by telomerase activity in a cell-cycle-dependent manner. It has been shown that the CST

complex is recruited to telomeres in S/G2 where it binds to the G-overhang, this terminates lengthening by telomerase and simulates C-strand synthesis by recruiting the Pol α primase (L. Y. Chen *et al.*, 2012). The understanding of this process has been further developed by functional studies of the CST subunits in telomerase positive colon cancer cells. Feng *et al* found that CTC-STN1 subunits associated with limiting telomerase action to prevent G-overhang overextension, while only STN1-TEN1 are required for synthesising of the C-strand (Feng *et al.*, 2018). STN1-TEN1 is a replication protein-A (RPA)-like complex and it is involved in preventing and rescuing stalled replication forks not only at telomeres but also at GC-reach regions genome-wide (Stewart *et al.*, 2012; F. Wang *et al.*, 2014; Hom and Wuttke, 2017).

1.2.4 Telomere replication

1.2.4.1 Telomere end replication problem

DNA replication is fundamental for cell division and many proteins participate in regulation of DNA replication in eukaryotes. Since DNA replication in eukaryotes is semi-conservative and all DNA polymerases can only syntheses DNA in the 5' → 3' direction, the DNA at the very end of the telomere cannot be fully synthesised. The incompletely replicated sequences at the end of linear chromosomes results in a gradual shortening of the telomere (Figure 1.2). The DNA polymerase epsilon (Pol ε) is required for leading strand synthesis of telomeric DNA using the C-reach strand as a template, while the DNA polymerase delta (Pol δ), and polymerase alpha primase (Pol α – primase) are required for lagging strand synthesis of telomeric DNA, using the G-reach strand as a template. Lagging strand synthesis uses RNA primers to initiate the DNA replication in discontinuous fragments (Okazaki fragments) (Okazaki *et al.*, 1968). The RNA primers are removed; the gaps between the Okazaki fragments are filled and ligated together but the removal of the terminal RNA primer leaves a gap resulting in incomplete synthesis of the lagging strand.. Thus at each round of cell division leaves 50-200 bp telomeric DNA replicated at the 3' end.
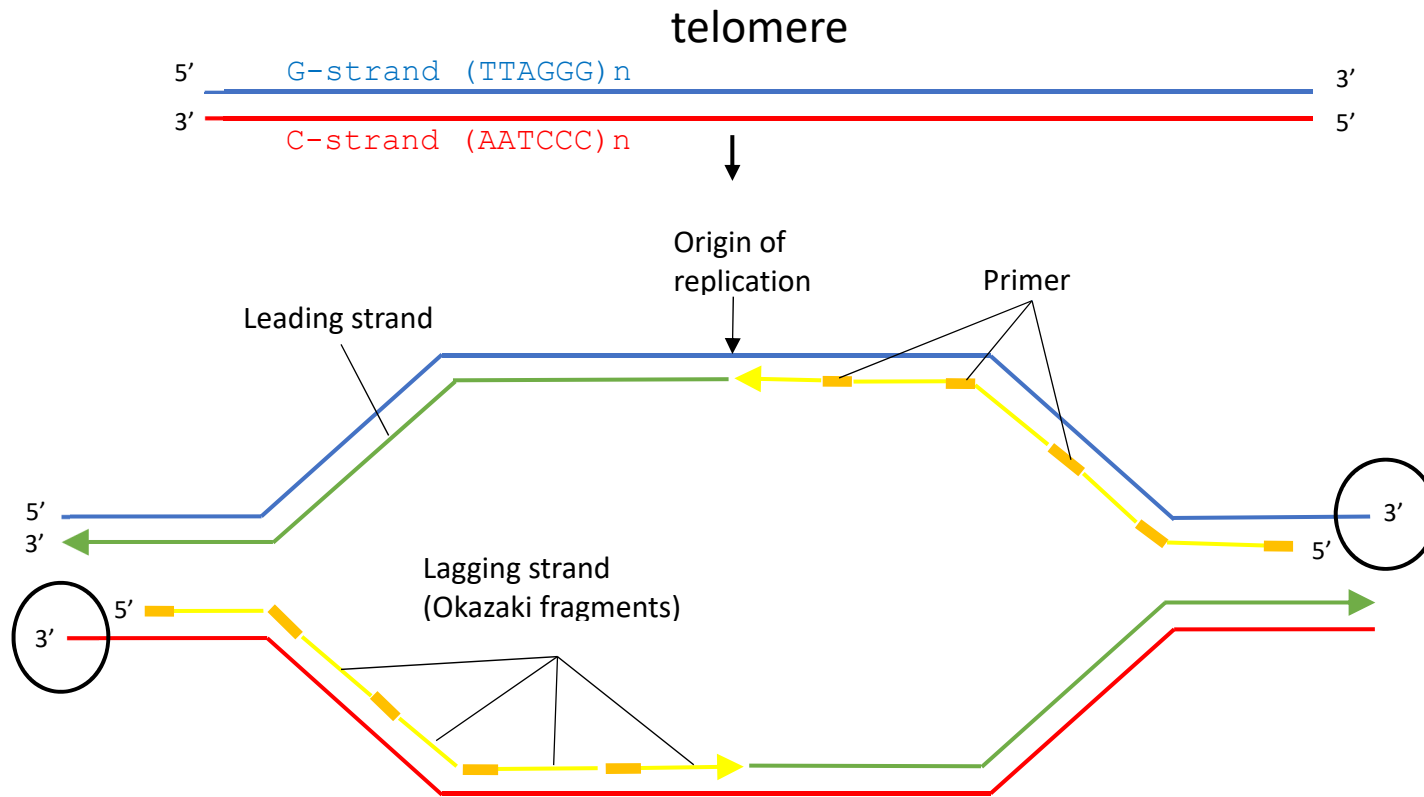
Figure 1.2 Model of telomere end replication problem. It illustrates discontinuous lagging strand synthesis primed by the Pol α-primase, which adds short RNA primers (orange bars). The continuous leading strand synthesis of telomeric DNA is mainly achieved by the activity of Pol ε. At the end of the telomere, the gap (black circles) after removal of the 5′ terminal RNA primer on the lagging strand cannot be filled in, and the telomere may become shorter with each successful round of replication.

1.2.4.2 Source of replication stress at telomeres

The repetitive nature of telomeric DNA results in telomeres manifesting as fragile sites during replication. This requires specialised helicases and nucleases to resolve the secondary structures at telomeres. There are three known source of replications stress at telomeres. First, the closed structure of the t-loop needs to be processed to linearise DNA for replication in S-phase. RTEL1 (Regulator of Telomere Elongation Helicase 1) which has strand displacement properties has been shown to be required for t-loop disassembly (Vannier *et al.*, 2012; Uringa *et al.*, 2012). The RecQ-like helicases which have 3′→5′-directed helicase activity include the Werner syndrome helicase (WRN) and Bloom syndrome helicase (BLM in human, Sgs1 in yeasts) both of which have been shown to interact with TRF2 to promote the unwinding of the t-loop in a cell-cycle-dependent manner (Opresko *et al.*, 2004; Lillard-Wetherell *et al.*, 2004). In the absence of RTEL1, a protein complex that includes SLX4 appear to cleave t-loops releasing extrachromosomal t-circles and resulting in negative regulation of telomere length (Vannier *et al.*, 2012) (Figure 1.3 A).

Secondly, G-quadruplex or G4 secondary structures (four guanines bound together by Hoogsteen hydrogen boding) may be formed on the G-reach strand of telomeric DNA during replication causing replication-fork arrest or stalling (Tarsounas and Tijsterman, 2013). The loading of RPA and POT1 onto the G-reach strand during replication is expected to reduce the occurrence of G4s (Zaug *et al.*, 2005; Salas *et al.*, 2006) and RECQ helicases such BLM (Sun *et al.*, 1998) and WRN (Mohaghegh *et al.*, 2001) have the capacity to unwind G4s at telomeres. FANCJ (Fanconi anemia (FA) complementation group J), another 5′–3′-directed helicase, also shows the capacity to unwind G4s in an RPA dependent manner (Y. Wu *et al.*, 2008). The DNA Replication Helicase/Nuclease 2 (DNA2) with 5′→3′ helicase and 3′- nuclease activity has the capacity to cleave G4s. Genetic knockout of *DNA2* in a mouse model resulted in telomere defects and genomic instability (Lin *et al.*, 2013). In additional to RTEL1's role in unwinding t-loops, it also appears to be able to resolve G4s so contributing to maintaining telomere stability (Vannier *et al.*, 2012). The ATRX (Alpha-Thalassemia/mental Retardation X-linked) is an ATP-dependent helicase, which belongs to the SWI/SNF

family of chromatin remodelling proteins. ATRX works with DAXX in maintaining the heterochromatin status at the repetitive regions by deposition of the histone H3.3 variant (McDowell *et al.*, 1999; Drane *et al.*, 2010; Goldberg *et al.*, 2010; Wong *et al.*, 2010; Lewis *et al.*, 2010). ATRX can bind G4s *in vitro* suggesting that it could also be involved in unwinding G4s (Law *et al.*, 2010).

The third potential source of replication stress at telomeres is the presence of TERRA (TElomeric Repeat-containing RNA) and R-loop formation. It was found that telomeres could be transcribed into long non-coding RNAs called TERRA. The initiation of TERRA transcription can occur in the subtelomeric regions internal to the telomere DNA and progresses through the telomeric DNA using the c-strand as a template (Azzalin *et al.*, 2007). TERRA forms RNA:DNA hybrids that results in TERRA R-loop (Rippe and Luke, 2015). It is thought that TERRA can be *cis*-acting (TERRA remains at the site where it was generated) or *trans*-acting (TERRA hybridises to DNA at another telomere) (Chu *et al.*, 2017). TERRA expression is cell-cycle regulated; it peaks at the G1/S transition and decreases in S /G2 in mammalian cells that use telomerase. TERRA levels are elevated in cell that use the ALT mechanism with accumulation in S/G2 associated with loss of ATRX (Flynn *et al.*, 2015) (Figure 1.3 B).

TERRA R-loop formation has been suggested to be a replication fork barrier that needs to be dissolved. RNA endonuclease H (RNase H) is the major enzyme involved in degrading TERRA and resolving the RNA:DNA loop in budding yeast (Balk *et al.*, 2013), and RNase H1 has been shown to regulate the TERRA levels in ALT+ cells (Arora *et al.*, 2014). Furthermore, a group of helicases and exonucleases are also predicted to dissolve R-loops consistent with their role in dissolving the DNA:RNA hybrids. Downregulation of UPF1 (Up-frameshift 1), a nucleic acid-dependent ATPase and 5′–3′ helicase activity, results in substantial increase in TERRA levels (Azzalin *et al.*, 2007). FEN1 (Flap Structure-Specific Endonuclease 1) has also been shown to have the capacity to remove RNA:DNA hybrids, suggesting it may have a role in removing the R-loops (Teasley *et al.*, 2015). PIF1 (Petite Integration Frequency 1 DNA Helicase) is another candidate as a TERRA regulator as exogenous expression of Pif1 in HT1080 cells results in significant reduction in telomerase activity and telomere shortening (Zhang *et al.*, 2006).
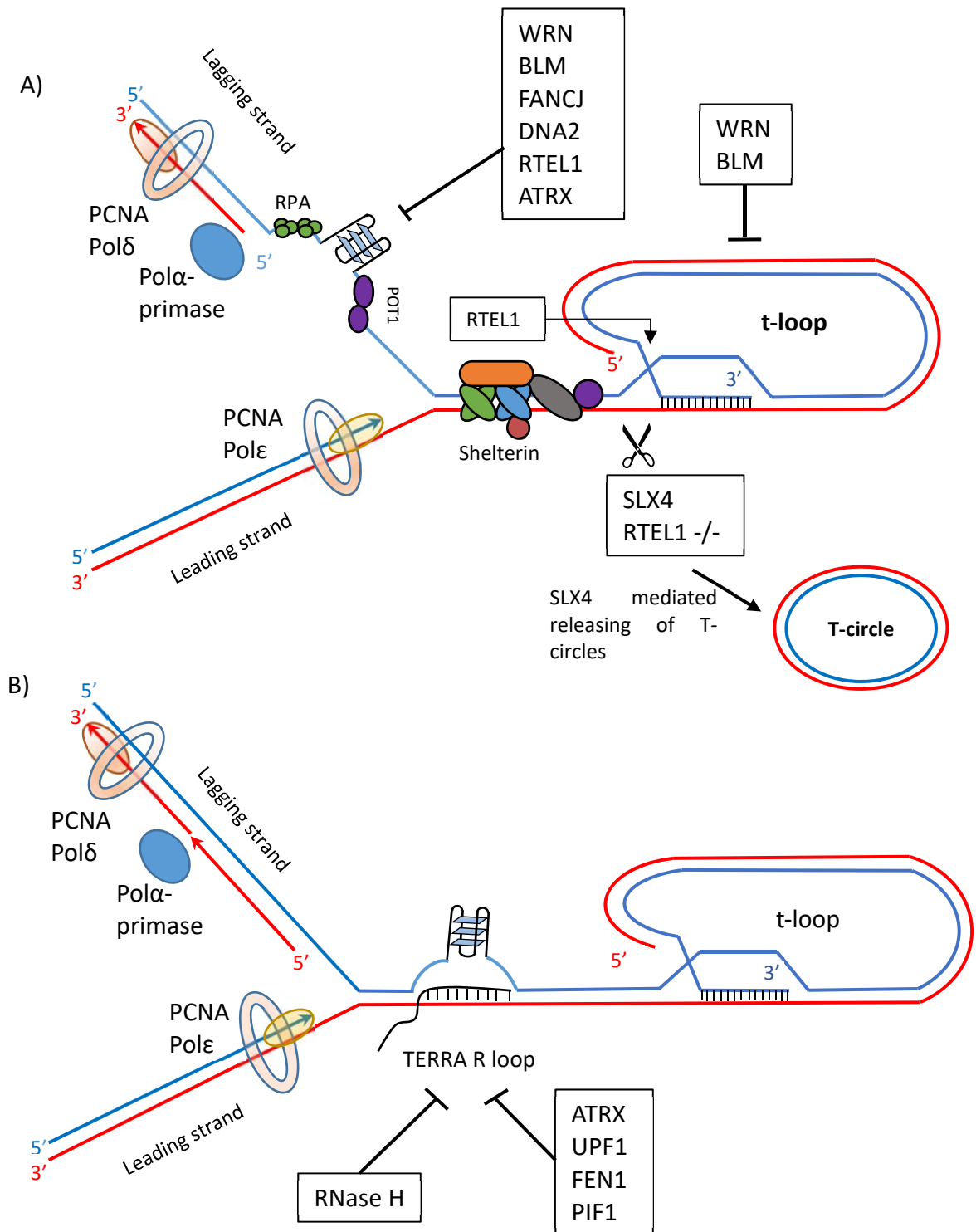
Figure 1.3 Sources of DNA replication stress at telomeres and their possible solutions. (A) Processing of G4s and t-loops to avoid replication fork stalling at telomeres. (B) Resolving or degrading TERRA R-loops.

1.3 Mechanisms of DSB repair

1.3.1 Classical non-homologous end joining (c-NHEJ)

The repair of DSB depends on when and where the break happens in humans. DSBs that occur throughout the cell cycle are repaired predominantly by the classical non-homologous DNA end joining (c-NHEJ) pathway. At the site of DSB, 53BP1 the DNA damage response factor and RIF1 (Rap1 interacting factor 1) promote c-NHEJ by suppression of the DSB end resection (Chapman *et al.*, 2013; Zimmermann and de Lange, 2014). The ends of DSB are recognised by the Ku70–Ku80 proteins that prevent the resection by MRN complex (MRE11-RAD50-NBS1, in humans; Mre11-Rad50-Xrs2 in yeast). Ku recruits the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). The DNA-PKcs together with Ku, forms the DNA-PK complex (Meek *et al.*, 2008) . NR4A1 interacts with DNA-PK to facilitate the DSB break repair by classical NHEJ (Malewicz *et al.*, 2011). This pathway also requires the ligase IV complex (LIG4/XRCC4/XLF) (Callebaut *et al.*, 2006; Bhargava *et al.*, 2018) and Pol μ and Pol λ to promote the ligation (Moon *et al.*, 2014) (Figure 1.4).

1.3.2 Alternative end joining (a-EJ)

In the absence of c-NHEJ, DSB ends are processed by the endonuclease MRN complexes to generate 15 to 100 nt 3' overhangs (Makharashvili *et al.*, 2014). The alternative end joining (a-EJ) also known as microhomology mediated end joining (MMEJ) shows to mediates the repair of DSBs by annealing of microhomolgy sequence ≤4 bp. This pathway involves LIG3, probably poly (ADP-ribose) polymerase 1 (PARP1), XRCC1 (Audebert *et al.*, 2004) , as well as Pol θ (Wood and Doublie, 2016). Furthermore, it showed that PARP2 involved in DSB repair pathway choice by PARP2 favours repair by limiting the access of 53BP1 at the DSB site and promoting efficient CtIP-dependent DNA end resection (Fouquin *et al.*, 2017) (Figure 1.4).
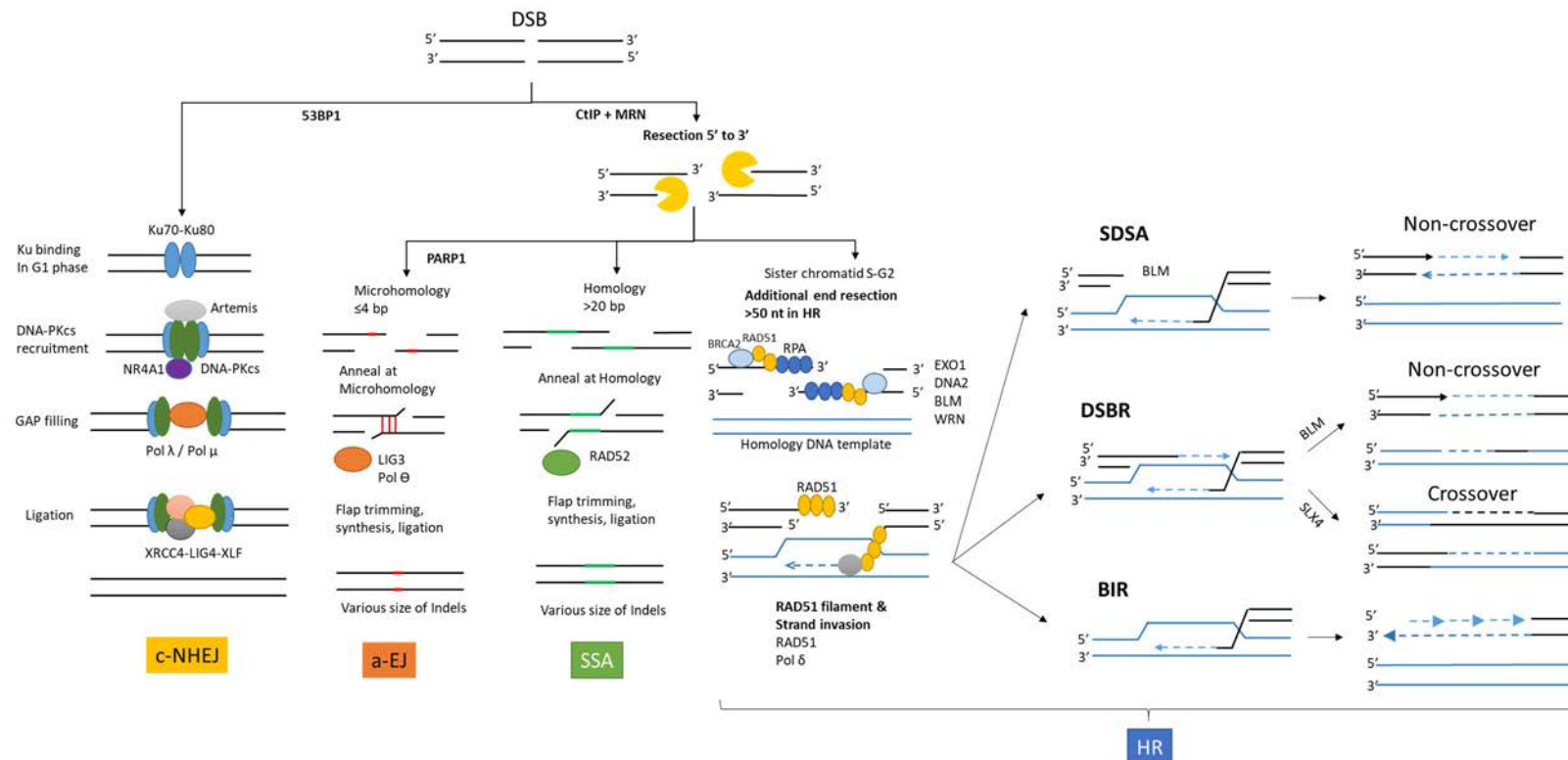
Figure 1.4 The pathway choice of double-strand break repair (DSB). The DSB can be processed by the classical non-homologous end joining (c-NHEJ) pathway, alternative end joining (a-EJ), or by homologous recombination (HR) pathways. c-NHEJ is positively regulated by the chromatin remodeller 53BP1 and Artemis when DSB ends are protected by Ku70-k80, and extensive end resection is not required. In contrast, the CtIP and MRN complex conduct 5' → 3' resection to generate 3' single-stranded DNA for a-EJ, SSA, or HR. In HR and SSA, BLM, EXO1, DNA2 and WRN may be required for more extensive resection.

1.3.3 Homology directed repair (HDR) pathways

Following the MRN resecting and particularly in S-G2 phases when the sister chromatid is present, ATM and CDK2 phosphorylates CtIP/ RBBP8 that thought to regulate the recruitment of extensive resection factors (Jazayeri *et al.*, 2006). The equipment of extensive resection factors such DNA2, BLM, WRN, and EXO1 generate long stretches of 3' overhangs, thereby committing the cells to homologous recombination (HR) or Single strand annealing (SSA). The RPA binds the long resected 3' single-stranded DNA strand and prevents spontaneous annealing between microhomologies (Deng *et al.*, 2014).

The SSA requires annealing at two homologous repeats in direct orientation repeat sequences near DSB with homology > 20 bp, no strand invasion and usually ends with large deletion (Bennardo *et al.*, 2008). It was shown that RAD52 (not RAD51) mediates the strand annealing of the two large homologous sequences (Kramer *et al.*, 1994). This pathway is mutagenic and results in large deletion events, but it can be use only when other HDR pathways are not available, specifically in the absence of RAD51 (Bhargava *et al.*, 2016) (Figure 1.4).

The HR implies strand invasion and therefore requires RAD51. In this pathway, BRCA2 binds to single strand overhang recruits RAD51 to single strand DNA promoting RPA displacement and RAD51 nucleoprotein filament formation (Carreira and Kowalczykowski, 2011). RAD51 nucleoprotein filament on single strand 3' overhang drives homology search and strand exchange with a homologous template (Sung *et al.*, 2003) (Figure 1.4). After the RAD51 nucleoprotein filament formation three possible subtype of HDR pathways can be used for repairing of DSB depends on the nature of DSB: synthesis-dependent strand annealing (SDSA), break-induced replication (BIR), or by double-strand break repair (DSBR) pathway. SDSA usually occurs where the invading strand is displaced from the donor molecule and annealed to the sequences on the other end of the break. Thus, SDSA usually ends with a non-crossover and no change introduced to the donor template (Miura *et al.*, 2012).

In DSBR, the second end of the break captured by the D-loop, this would result in a double Holliday junction intermediate, which can then be resolved by endonucleases and helicases. BLM-TOP3A-RMI1/2 complex shows to play role in strand migration and dissolving the HJs in non-crossover products (S. H. Chen *et al.*, 2014). By contrast SLX4–SLX1–ERCC4 complex can resolve the HJ in crossover protects (Matos and West, 2014; Sobinoff *et al.*, 2017). Mus81 endonuclease was also showed to mediate the cleavage of the HJs in yeasts (Boddy *et al.*, 2001), and human cells (X. B. Chen *et al.*, 2001).

In contrast to other HR pathways, the BIR is used only in cases where a single end is present. BIR contributes to the repair of broken replication forks and maintaining telomeres. BIR is a conservative repair pathway. In BIR, the invading strand is extended conservatively and the newly synthesised DNA is displaced from the donor to serve as a template for synthesis of the second strand as far as the bubble migrated. However, resolving the HJ and replication fork formation promotes the semi-conservative BIR repair (Donnianni and Symington, 2013; Saini *et al.*, 2013).

BIR has been described as a conserved pathway that plays a key role across different organisms. BIR is used by bacteria and viruses to solve the replication problem. In *E. coli,* it is used to recover broken replication forks during the replication of the circular chromosomes (Marians, 2000). BIR is not required in eukaryotes for completion of DNA replication as they rely on telomerase to solve the end replication problem. However, it has been shown that it is required for telomere elongation in yeast that lack telomerase (McEachern and Haber, 2006). Similar roles for BIR have been seen in human cancer cells. BIR can be used to restore the collapsed replication forks, which results in genomic rearrangements and genetic mutations (Costantino *et al.*, 2014). BIR is also likely to be responsible for the ALT mechanism in cancer cells (Royle *et al.*, 2009; Mendez-Bermudez *et al.*, 2012; Sakofsky and Malkova, 2017).

1.4 Telomere Maintenance Mechanisms (TMM)

1.4.1 Immortalisation of human cells by expression of SV40- Large T antigen

Telomere length can predict the replicative capacity of primary cells in culture (Allsopp *et al.*, 1992). The life span of normal human newborn or foetal diploid fibroblasts in culture was shown by Hayflick (1965) to be 50-60 population doublings (PDs), until they reached a state of non-proliferation called cellular senescence, also known as the mortality stage 1 (M1). Replicative senescence is characterised by growth arrest in the G1/S phase of the cell cycle (Goldstein, 1990). Expression of the large T antigen from the Simian virus 40 (SV40-LT), using mammalian expression plasmid, in human fibroblasts has been shown to extend lifespan of cells beyond the normal M1 checkpoints (Wright *et al.*, 1989) (Figure 1.5).

The SV40 large T antigen has numerous targets in human cells. Most importantly, SV40-LT binds to and inactivates p53 and Rb proteins, the key regulators of the G1/S transition. SV40-LT inhibits p53 activity by forming transcriptionally inactive LT-p53 complex (Bargonetti *et al.*, 1992; Sheppard *et al.*, 1999; Xing *et al.*, 2001). Another study showed that SV40-LT binds to and inhibits three Rb family proteins (pRb, p107, and p130) forming a LT-Rb complex, and consequently increasing the transcriptional activity of E2F which leads to a positive regulation of the cell cycle and an increase in the proliferation rate (Rathi *et al.*, 2009). Stable expression of SV40-LT is also associated with changes in cell morphology, and loss of cell-cell contact inhibition (Vessey and Culp, 1978; Porras *et al.*, 1999). However, pre-crisis SV40-LT transformed cells are not immortalised. In the absence of functional p53/Rb, the SV40-transformed cells can continue to divide until they hit the second replication barrier known as crisis or mortality stage 2 (M2) (Wright and Shay, 1992; Shay and Wright, 2005). Crisis is characterised by short and dysfunctional telomeres that trigger Breakage-Fusion-Bridge cycles leading to chromosome instability (Wright and Shay, 1992; Counter *et al.*, 1992; Shay *et al.*, 1993; Shay and Wright, 2005). Populations of cells that enter telomere driven crisis show a reduction in growth rate and a substantial increase in the frequency of cell death. Very few cells escape crisis and become immortalised by activating a Telomere Maintenance Mechanism (TMM): The majority of human cancers (~85%) activate telomerase, while a

significant minority of cancers (~15%) activate the ALT recombination-based mechanism (Henson *et al.*, 2002). The immortalisation frequency among SV40-LT expressing cells is a low (~ 1 x 10^{-7} in human fibroblasts, and 1 x 10^{-5} in epithelial cells) (Shay *et al.*, 1993). The cells that bypassed crisis are expected to continue to express SV40-LT to keep M1 inactivated (Wright *et al.*, 1989). The probability of activation either TMM appears to be influenced by the cell type and tissue origin. Cells with a mesenchymal origin (such as fibroblasts) appear to activate the ALT mechanism nearly as frequently as they activate telomerase, whereas human cancers with an epithelial origin (such as carcinomas) preferentially activate telomerase (Colgin and Reddel, 1999; Heaphy *et al.*, 2011).
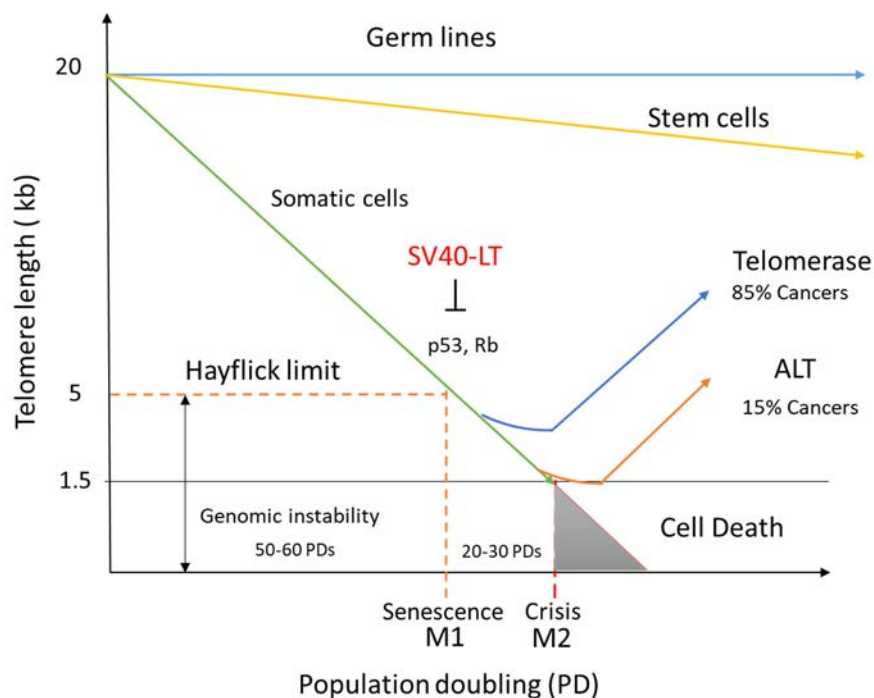
Figure 1.5 Human fibroblast immortalisation mediated by SV40-transformation. The diagram shows the life span of normal and SV40-transformed fibroblasts. The normal human fibroblast hit senescence (M1 stage) after 50-60 PDs, while the SV40-large T antigen expressing cells have an extended lifespan and continue to grow beyond the normal M1 limit until they are reach crisis (M2 stage), when the majority of cells die. The few cells that emerged from the crisis have the potentiality for unlimited cell division.

1.4.2 Telomere maintenance by telomerase

1.4.2.1 Mechanism of telomere elongation by telomerase

Telomerase activity is not detectable in normal somatic cells where these cells have a limited proliferative capacity (Shay and Wright, 2000), but most cancer cells express telomerase (Henson *et al.*, 2002). The human telomerase complex consists of telomerase reverse transcriptase TERT (human homolog of yeast Est2 in *Saccharomyces cerevisiae*), telomerase RNA (TERC) which contains '5- CAAUCCCAAUC-3' as a template for reverse transcription into TTAGGG DNA, and four other proteins (Dyskerin, NOP10, NHP2, and GAR) (Greider, 1996; Sandin and Rhodes, 2014) (Figure 1.6). Telomerase elongates telomeres by a highly regulated mechanism in which telomerase is recruited to the G-overhang at the end of the telomere by interacting with TPP1 and via base pairing of the TERC template region with the terminal bases of the G-overhang. Once the interaction between telomerase and telomere is established, the elongation process can start by reverse transcription of the RNA template region, followed by translocation of telomerase the end or the newly synthesised repeat unit. Repetition of the elongation-translocation steps results in addition of more TTAGGG repeats (Sandin and Rhodes, 2014).

1.4.2.2 Regulation of telomerase expression

Telomerase activity is tightly regulated in human cells via expression of the *TERT* gene. Expression of *TERT* is repressed in most somatic cells but it is upregulated in cancer cells (Cong *et al.*, 1999; Takakura *et al.*, 1999). The *TERT* gene is located on chromosome 5 and consists of 16 exons, spanning ~35 kb. The *TERT* promoter is GC-reach and lacks TATA and CAAT boxes; however, it contains many sites for several transcription factors showing a high level of regulation by multiple factors (Cong *et al.*, 1999). Several proto-oncogene transcription factors can activate *TERT* expression including c-Myc, Sp1, and HIF-1 but the activity of *TERT* can be supressed by tumour suppressors such as p53, Rb, MEN1 and WT1 (Kyo *et al.*, 2008; Cukusic *et al.*, 2008). DNA hypomethylation around the transcription start site of the *TERT* promoter have been shown to facilitate *TERT* transcription (Kyo *et al.*, 2008).
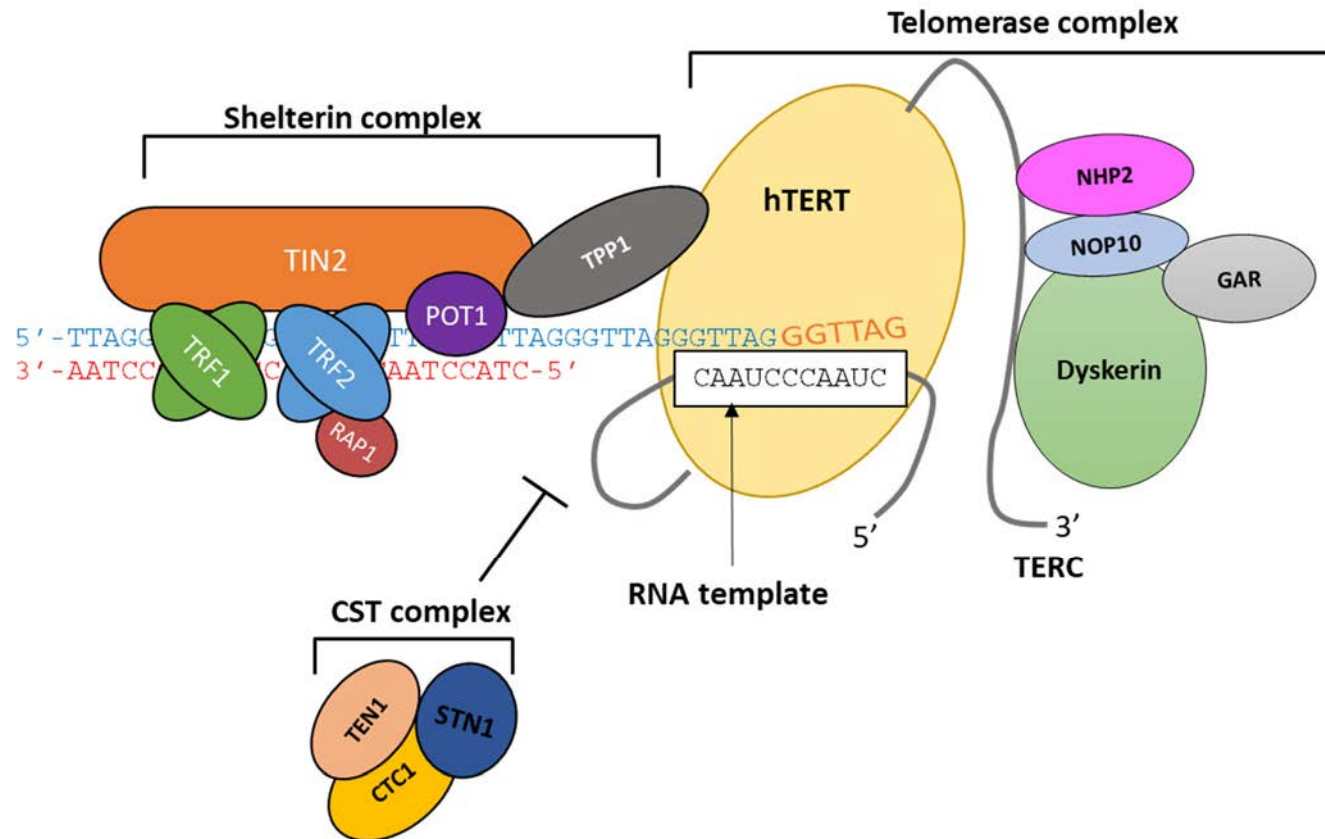
Figure 1.6. Telomere lengthening by telomerase. The diagram shows the telomerase, including TERT and other subunits binding to the G-overhang of the telomere and synthesizing telomeric DNA de novo using telomerase RNA (TERC) which contains '5-CAAUCCCAAUC-3' as a template. The CST complex works as negative regulator for telomerase activity

1.4.3 Alternative Lengthening of Telomeres (ALT) Mechanism

1.4.3.1 Alternative lengthening of telomeres in yeast

In 1984, two different mechanisms of telomere lengthening were proposed in budding yeast. Yeast cells expressing Est1 (Telomere elongation protein in yeast) maintain their telomeres by de novo addition of telomere repeats (Shampay *et al.*, 1984), which is similar to human cells that use telomerase. Whereas the yeast strains that lack the EST1 gene showed progressive telomere shortening until they reach the crisis where the majority of cells are dead. However, a proportion of cells are able to overcome the estl-Δ mutation and showed an increase in recombination events at the subtelomeric regions (Lundblad and Blackburn, 1993). A similar observation to the est1 mutation was found in stains that lack other genes that encode telomerase subunits in yeast such as EST2 (Telomerase reverse transcriptase protein in yeast) (Lendvay *et al.*, 1996). This mechanism underlies the telomere extension in survivors known as alternative lengthening of telomeres (ALT).

The telomerase-negative crisis survivors have been classified according to their telomere composition into type I and type II survivors. The type I survivors show duplication of tandem Y elements with short telomeres. In contrast, the type II survivors show long and heterogeneous telomeres. However, the ALT mechanism in both type I and type II survivors is dependent on Rad52; an essential homologous recombination protein (Teng and Zakian, 1999). The proposed mechanism of ALT in type II survivors suggests a recombination-based event that mediates telomere re-lengthening in these cells. Re-expressing telomerase results in type II survivors restoring the length of short telomeres, while graduate shortening in length of extra-ordinary long telomeres has observed (Teng and Zakian, 1999).

The ALT in yeast have been shown to rely on BIR for telomere elongation. It was proposed that the critical length of telomeres impaired the binding of associated proteins, where the 3' tail can invade a long telomere of another chromosome, resulting in a net increase in telomere length after a successful telomere recombination (McEachern and Haber, 2006).

1.4.3.2 Alternative lengthening of telomeres in human

1.4.3.3 ALT markers in human cells

In 1995, long heterogeneous telomeres were identified in human *in vitro* immortalised cells that lack telomerase expression (Bryan *et al.*, 1995), and in a number of tumours (Yeager *et al.*, 1999). Telomere length in human ALT+ cells ranges from 2 to 50 kb, and rapid changes in telomere length have been detected over just one or a few cell divisions. In contrast, human cells with telomerase activity show a stable telomere length with a limited length distribution (Murnane *et al.*, 1994; Bryan *et al.*, 1995).

PML nuclear bodies are structures containing PML and Sp100 proteins that bring large sets of nuclear proteins into close proximity. PML nuclear bodies can serve as a platform for multiple nuclear functions (Borden, 2002). They also vary in size and number between cells. Small PML bodies are usually found in ALT negative cells such HT1080. The small PML bodies often contain ATRX, DAXX, BRCA1, HP1, SP100 proteins. Thus, it has been suggested that small PML bodies in ALT-negative cells might participate in resolving secondary structures formed during replication of repetitive DNA and they may be involve in the condensation and maintenance of heterochromatin in these cells (Luciani *et al.*, 2006; Draskovic *et al.*, 2009).

ALT+ cell lines and tumours often contain specialised form of PML Nuclear Bodies called ALT-associated PML bodies (APBs) (Henson *et al.*, 2005). These APBs contain a large quantity of telomeric DNA that can be chromosomal or extrachromosomal. They also contain components of the Shelterin complex (Yeager *et al.*, 1999; Tarsounas *et al.*, 2004; Nabetani *et al.*, 2004; Silverman *et al.*, 2004; Draskovic *et al.*, 2009; Lang *et al.*, 2010). Apart from telomere binding proteins, APBs contain proteins involved in DNA repair and homologous recombination such as RAD52 and RAD51 (Yeager *et al.*, 1999), WRN, BLM (Stavropoulos *et al.*, 2002), SUMOylation SMC5/6 complex (Potts and Yu, 2007), MRN (MREl 1, RAD50, NBS1) (Borde and Cobb, 2009; G. Wu *et al.*, 2000).

APBs are associated with cell cycle arrest in G2/M phage. One reason may be the accumulation of single-stranded DNA in APBs as seen following depletion of RPA

protein in U2OS and GM-847 ALT+ cell lines (Grudic *et al.*, 2007). Furthermore, it has been proposed APBs function as platform for telomere length extension by ALT (Conomos *et al.*, 2012; Conomos *et al.*, 2014). Alternatively, APBs could serve as a repository for extrachromosomal telomeric repeat (ECTR) DNA that can be linear or circular including double-stranded t-circles and single-stranded or partially double stranded telomeric C- and G- circles. It has been speculated that ECTR is a by-product of recombination events and releasing of t-loops from ALT+ telomeres (Figure 1.7) (Cesare and Griffith, 2004; Fasching *et al.*, 2007). T-circles are not specific to ALT+ cells as they are also present in normal somatic cells, but C-circles (and the less abundant G-circles) are regarded as specific markers for ALT activity (R. C. Wang *et al.*, 2004; Henson *et al.*, 2009; Henson and Reddel, 2010).

C-circles can be partially double-stranded circular molecules of telomeric DNA with a closed circular C-strand that have the potential to serve as a template for rolling circle amplification and telomere elongation in ALT+ cells (Grudic *et al.*, 2007; Henson *et al.*, 2009). The abundance of C-circles in ALT+ cells is 1000 times higher than ALT- cells and this level changes dynamically in response to ALT activation or inhibition. Expressing SP100 (an ALT suppressor) in IIICF/c ALT+ cells resulted in a rapid reduction in the abundance of C-circles. Thus, the C-circles assay is regarded as a sensitive and quantifiable marker for ALT (Henson *et al.*, 2009).

T-circles

= G-strand
= C-strand

Potential types of C-circles

Potential types of G-circles
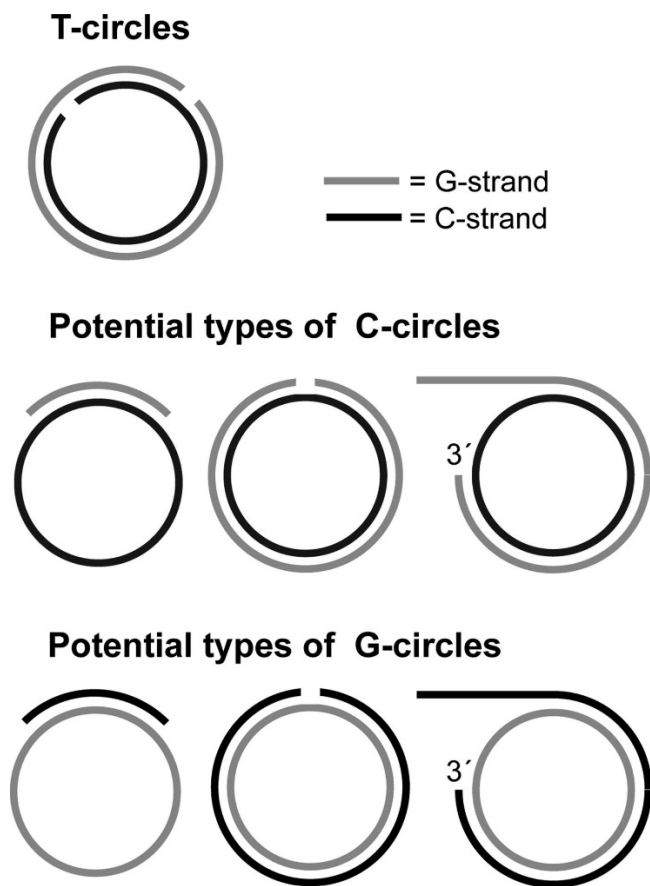
Figure 1.7. Extrachromosomal telomeric circles. This figure illustrates the three types of extrachromosomal telomeric circles in ALT+ cells: double stranded t-circles, which are unlikely to be templates for rolling circle amplification, while C-circles and G-circles could be templates for auto-priming rolling circle reactions. (Henson and Reddel, 2010)

1.4.3.4 Telomeric recombination in human ALT cells

An elevated level of sister-chromatid exchange at telomeres is a marker for increase of post-replication recombinations in ALT+ cells. Assaying for T-SCEs is carried out by using chromosome orientation fluorescence *in situ* hybridization (CO-FISH) (Londono-Vallejo et al., 2004).

The earliest studies of the ALT mechanism in human cells used plasmids integrated into telomeres or adjacent sequences as tags. If the tag was integrated within telomeric DNA, it could be copied from one telomere to another, whereas if the tag was located in the subtelomeric region, no copying of the tag was detected (Dunham *et al.*, 2000). This resulted in the proposal that the ALT mechanism involved copying between telomeres on different chromosomes known as "inter-telomeric recombination". These recombination-like events have also been characterised by studying classes of mutations that arise in the variant repeat regions of telomeres in ALT+ cells but not in telomerase positive cells (Varley *et al.*, 2002). ALT+ cells also show a remarkably high level of instability at the MS32 minisatellite locus, but not other GC-reach minisatellites. The mutation rate at MS32 varies between ALT+ cell lines and tumours but the relationship to ALT activity is not understood (Royle et al., 2009).

The strand invasion in ALT can be achieved by either BIR or SDSA in the presence of sister chromatid telomere, or by BIR between non-homologous telomeres or between telomeres and extra-chromosomal telomeric repeats (Varley *et al.*, 2002; Royle *et al.*, 2009; Mendez-Bermudez *et al.*, 2012). The possible models includes intra-telomeric strand invasion of sister-chromatids of the same chromosome, while an inter-telomeric models involves strand invasion of a telomere on another chromosome or extrachromosomal linear or circular telomeric DNA (Dunham *et al.*, 2000; Varley *et al.*, 2002; Cesare and Reddel, 2010) (Figure 1.8).
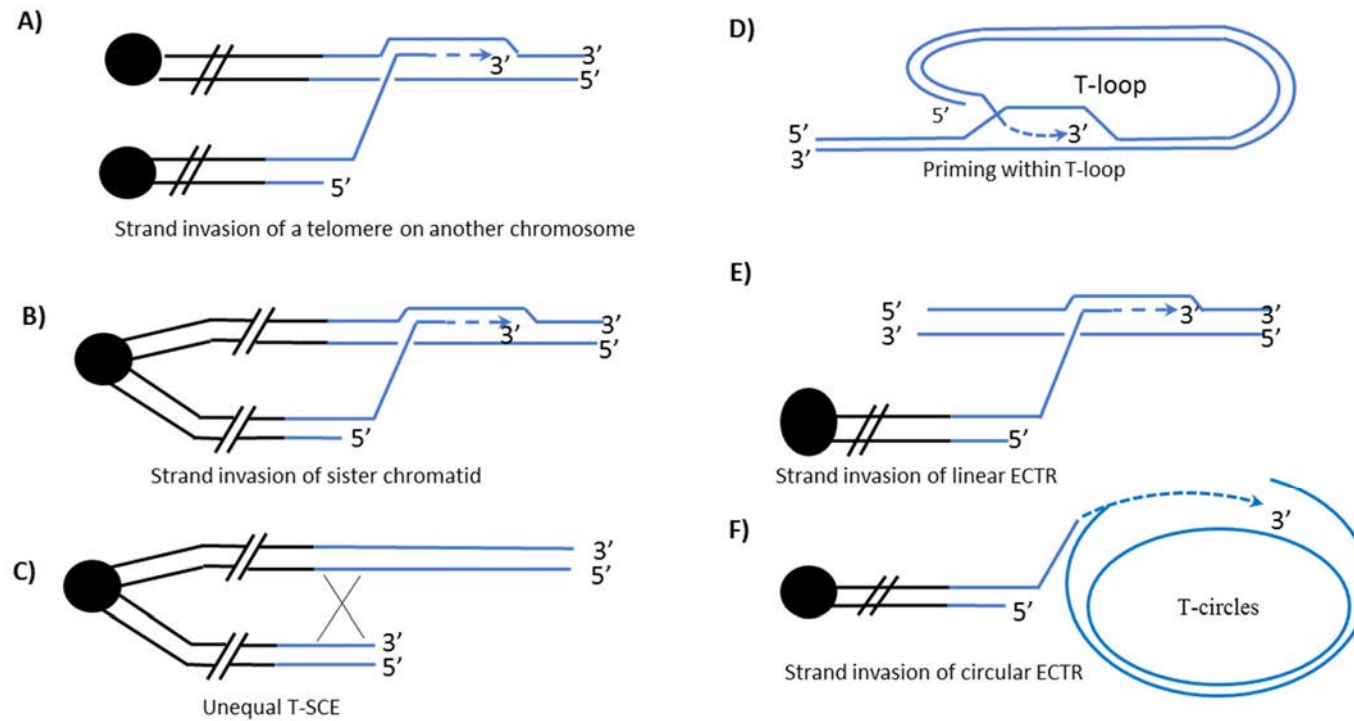
Figure 1.8 Models of homologous recombination-based telomere DNA replication in cells that use ALT. The 3' end of the telomere invades a double stranded DNA, which is used as a template for telomere elongation. The template could be another telomere (A) the sister-telomere (B), unequal telomere sister-chromatid exchange (T-SCE) although this does not result in a net gain of telomeric DNA (C), Priming and rolling circle replication within t-loops (D), or extrachromosomal DNA either in linear (E) or circular form (F). Figure modified from (Cesare and Reddel, 2010)

1.4.3.5 Regulation of ALT in human cells

Since the ALT mechanism was first described significant advances have been made in understanding the underlying mechanism and in the relationship to DDR at telomeres and many genes involved have been identified (see list of ALT related genes in Appendix 1.1). Factors that promote and suppress ALT are discussed in this section.

1.4.3.5.1 Replication stress in ALT

Although the initial events that trigger telomere elongation by ALT are unclear, several studies suggested that ongoing replication stress in G-reach sequences can result in replication fork stalling and double strand break formation, which can be restarted by BIR repair pathway (Figure 1.9).

ATRX works with its partner DAXX, as a chaperone complex for the histone variant H3.3, in order to maintain heterochromatin in some of the repetitive regions of the genome. These repetitive regions include telomeres and pericentromeric regions (McDowell et al., 1999; Drane et al., 2010; Goldberg et al., 2010; Wong et al., 2010; Lewis et al., 2010). Loss of ATRX or its partner DAXX may be a source of replication stress and G4 formation. Exogenous expression of ATRX in ALT+ cells resulted in a reduction in replication fork stalling by resolving G4 and reduction in all ALT markers in DAXX/H3.3 dependent manner (Clynes *et al.*, 2015). The role of ATRX in ALT has been discussed further in chapter 5.

ALT telomeres are characterised by a reduction in the levels of H3K9me3 (trimethylation of lysine 9 on histone H3 protein subunit) and this correlates with elevated levels of TERRA, while the opposite occurs at the telomeres of cells with telomerase activity (Episkopou *et al.*, 2014; O'Sullivan and Almouzni, 2014). R-loop formation, which may result from the elevated level of TERRA in ALT+ cells (Flynn *et al.*, 2015), may also participate in replication stalling in telomeres of ALT+ cells.
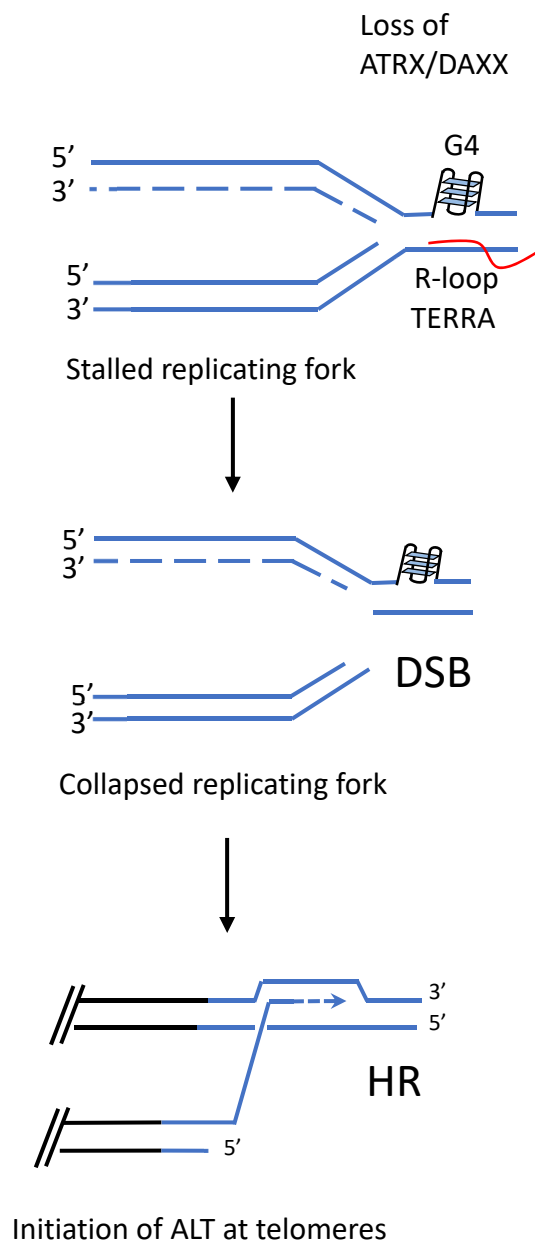
Figure 1.9 Model of replication stress at telomeres in ALT+ cells. Loss of ATRX/DAXX may be associated with increased G4 formation in the G-reach strand of telomere during replication. The elevated level of TERRA in ALT+ cells may increase R-loops (RNA:DNA hybirds) at telomeres. Both of G4s and R-loops could increase the risk of the replication fork stalling resulting in DSB formation, which may trigger ALT.

1.4.3.5.2 Telomere capping in ALT

TRF2 and POT1 are essential proteins for telomere capping and they repress the DNA damage signalling of ATM kinase signalling and ATR kinase signalling (Denchi and de Lange, 2007; Celli and de Lange, 2005). Critically short telomeres have reduced shelterin levels, which result in the formation of TIFs at these telomeres. Nuclei of ALT+ cells show many of TIFs but not all the telomeres are short, which may indicate that the telomeres in ALT+ cells are partially uncapped and this facilitates homologous recombination-like activity.

Two members of the Structural Maintenance of Chromosome 5 and 6 (SMC5/6) protein complexes are associated with a SUMO ligase protein called hMMS2. The hMMS21 ligase of the SMC5/6 complex SUMOylates multiple telomere-binding proteins, including TRF1 and TRF2. Depletion of SMC5/6 protein in ALT+ cells results in shortening of telomeres and inhibition of cell growth (Potts and Yu, 2007). Furthermore, SMC5/6 complexes are implicated in restarting collapsed replication forks by resolving HR structures during the repair of DSB by HR between sister-chromatids. They may also work with endonucleases, such as MUS81 and MMS4, in resolving Holliday junctions (Kegel and Sjogren, 2010). This suggests that they might play an important role in the synthesis of long tracks of telomeric DNA.

Fanconi Anaemia (FA) proteins such as FANCA, FANCL, BRIP1/ FANCJ and FANCD2 are involved in the recognition and repair of damaged DNA and recombination repair of DSBs. Some FA proteins show localisation in APBs and it has been shown that the monoubiquitination activity of FANCA and FANCL are required for recruitment of FANCD2 to APBs. Depletion of either FANCD2 or FANCA expression was implicated in the loss of telomeric DNA and reduction in T-SCEs in ALT+ cell lines (Fan *et al.*, 2009).

The CST complex binds to the 3' single-stranded overhang of telomeres, via the CTC1 protein, and this may compete with binding by POT1-TPP1. A recent study showed that the CST complex may also play a role in ALT. Tracking Flag-tagged-CTC1 in U2OS cells using FACS analysis of cell cycle showed that CST is associated with telomeres in APBs throughout the cell cycle. Furthermore, depletion of STN1 or CTC1 resulted in

reduction of C-circles and t-circles, while complete repression of STN1 effected cell proliferation, suggesting a role for CST complex in regulating ALT (Huang *et al.*, 2017).

1.4.3.5.3 Processing of DSB by ALT

Consistent with ALT being a HR-like mechanism, the MRN complex is required for ALT. MRN carries out 5' →3' resection of DSBs generating 3' single strand DNA overhangs (Bernstein and Rothstein, 2009), as an intermediate step for DSB resection by one of HR pathways. MRN is found in APBs (Zhong et al., 2007, Henson et al., 2009). Depletion of Mre11 results in telomere erosion in some ALT+ cell lines, while down-regulation of Nbs1 repression results in a reduction in t-circles (Zhong *et al.*, 2007). EXO1 may play an additional role in 5' →3' resection generating longer 3' overhang for HR mediated DSB repair (Bernstein and Rothstein, 2009). BLM, WRN have also been shown to take part in 5' →3' resection by interacting with DNA2 (Sturzenegger *et al.*, 2014).

Sp100, a major component of the PML bodies, is known as a negative regulator of MRN. Overexpression of Sp100 resulted in sequestering the MRN complex away from APBs in ALT+ cells and this resulted in repression of the ALT mechanism, evidenced by progressive telomere shortening at 121 bp per population doubling of IIIcfa ALT+ cells (Jiang *et al.*, 2005).

Following resection at DSBs, the RPA protein temporarily coats the 3' single strand overhang. The isoforms of the RAD51 recombinase protein subsequently replace the RPA prior to the strand invasion into a template molecule. Expression of a TRF1–FokI fusion protein in U2OS cells showed that RAD51 mediated telomere clustering (Cho *et al.*, 2014). In addition, the HOP2-MND1 heterodimer of meiotic proteins were also shown to associate with RAD51 at the site of damaged telomeres in ALT+ cells. Downregulation of HOP2 and MND1 resulted in resolving of telomere clusters in APBs. This suggests that RAD51 can work with the HOP2-MND1 heterodimer to mediate the inter-telomeric strand invasion and copying (Bugreev *et al.*, 2014; Cho *et al.*, 2014). RAD52 may also be required for recombination-like activity at telomeres in ALT+ cells.

After the strand invasion of a telomeric template, DNA polymerase delta subunit 3 (POLD3) has been shown to be recruited to the site of break-induced DNA replication in U2OS cells (Costantino *et al.*, 2014). Downregulation of POLD3 results in reduction of telomere length, and in the abundance of C-circles in U2OS cells (Dilley *et al.*, 2016). BLM plays important roles in unwinding DNA structures that resemble Holliday junctions and DNA secondary structures like G4s (Mohaghegh *et al.*, 2001; van Brabant *et al.*, 2000). BLM deficient cells are characterized by chromosomal abnormalities and an elevated level of sister chromatid exchanges (Bhattacharyya *et al.*, 2009) and depletion of BLM in ALT+ cells resulted in telomere shortening and growth suppression (Bhattacharyya *et al.*, 2009). It has also been shown that BLM works as a part of a complex (BLM-TOP3A-RMI1/2) to promote strand migration and long tract telomere extension followed by dissolution of double Holliday junction in U2OS ALT+ cells. The BLM-TOP3A-RMI1/2 complex requires POLD3 and RAD51 to promote telomere synthesis in ALT+ cells. In contrast, this process is abolished by the SLX4–SLX1–ERCC4 complex, which promotes resolution of the HJs in crossover and no telomere extension (Sobinoff *et al.*, 2017).

1.5 Orphan Nuclear Receptors

Orphan nuclear receptors (ORs) were grouped together as nuclear receptor families for which the endogenous ligands had not been identified. Human ORs play essential roles in embryonic development, angiogenesis, metabolic homeostasis, but also in progression of various diseases including cancer (Robinson-Rechavi *et al.*, 2003; Safe *et al.*, 2014).

The phylogenetic analysis of nuclear receptors families showed that these ORs belong to six nuclear receptors families (NR0, NR1, NR2, NR3, NR4, and NR6), and each gene family has diverged separately from its proposed common ancestor NR2A1 (HNF4-alpha) (Bridgham *et al.*, 2010) as shown in Figure 1.10.There is another group of nuclear receptors known as adopted orphan receptors. The members of this group were originally classified as ORs, but recently their ligands have been identified such as NR2A1/2 (Hepatocyte nuclear factor-4 alpha/beta (HNF4 α/β)), NR5A1 (Steroidogenic factor 1 (SF1)), and NR5A2 (Liver receptor homolog-1 (LRH-1)) (Safe *et al.*, 2014; Shi, 2007; Giguere, 1999).
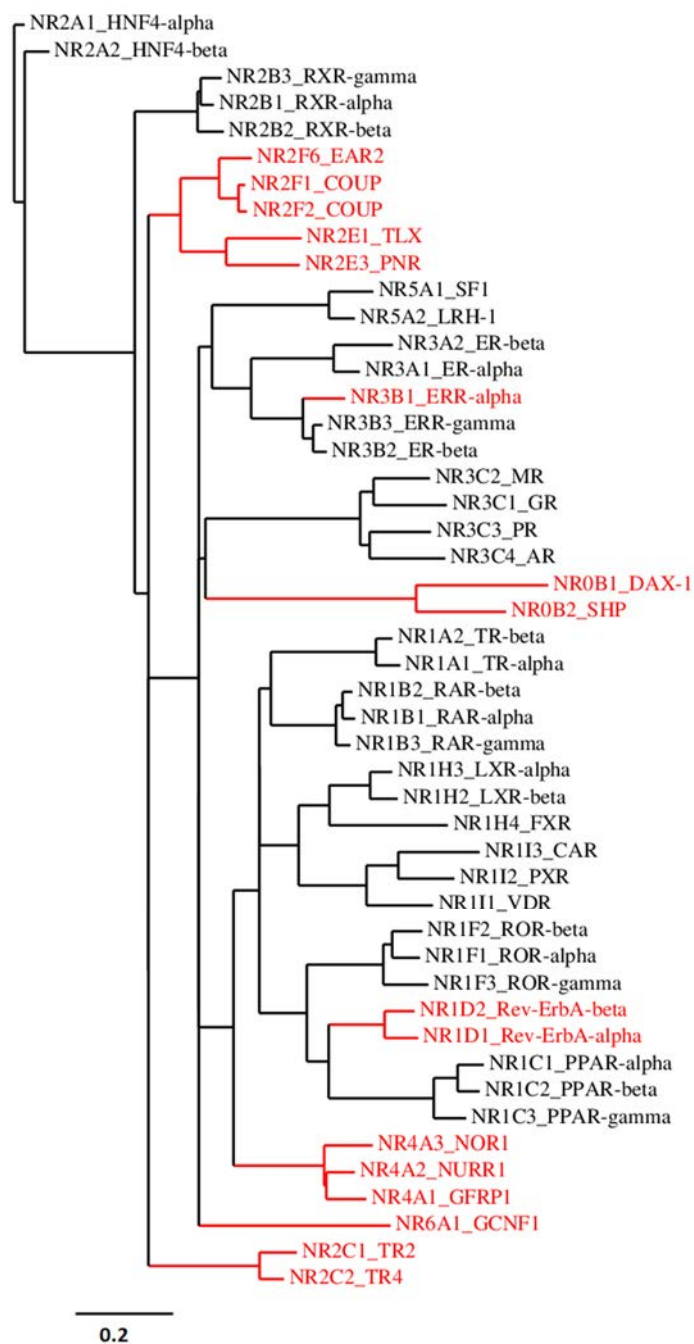
Figure 1.10 Phylogenetic tree of human nuclear receptors. A well-resolved phylogenetic tree was constructed by the maximum likelihood method and protein sequence of 48 human nuclear receptors downloaded from Swissport bank on 01/08/2017. NR2A1 (HNF4-alpha), the closest NR for their proposed ancestor (Bridgham *et al.*, 2010), was used as the outgroup to demonstrate the evolutionary relationship among the members of nuclear receptors superfamily. Red coloured, orphan nuclear receptors. The scale indicates the number of amino acid substitutions per site.

1.5.1 Domain structure and DNA binding by ORs

The classic structure of orphan nuclear receptors share a common molecular structure, which consists of four functional domain regions: an Activation Function 1 (AF1) domain, DNA-binding domain (DBD), hinge domain, and Ligand-Binding Domain (LBD). AF1 is an extremely variable domain that facilitates ligand-independent transcriptional activation of ORs. The DBD is a highly conserved domain in ORs, which consists of two cysteine reach zinc fingers responsible for DNA binding specificity and dimerization. The LBD is moderately conserved in ORs and plays roles in ligand binding, dimerization and interacts with co-activators. The hinge domain is a flexible region in ORs that links the DBD with LBD and play roles in post-transcriptional targeting of ORs (Sentis *et al.*, 2005; Giguere, 1999; Claessens and Gewirth, 2004; Marciano *et al.*, 2014; Chandra *et al.*, 2008). The structure analysis of 16 orphan receptors shows high conservation at the DBD, less conservation at the LBD domain and a significant level of diversity among the AF1 and hinge domains. However, atypical structure of orphan nuclear receptors belonging to the NR0B family shows the AF1, DBD, hinge domains are replaced by a 65 to 70 amino acid motif rich in alanine and glycine (Martin and Tremblay, 2010) (Figure 1.11).

The crystal structure studies of OR ligand-binding domains have revealed a variation in the size of LBDs pockets that range from a totally sealed cavity (as in NR4A2), a small LBD pocket (in NR2E3) to relatively large pockets (as in NR0B2) (Z. Wang *et al.*, 2003; Sablin *et al.*, 2008; Tan *et al.*, 2013). This suggests that ligand-mediated activity may not exist for all ORs, and the transcriptional activity of such ORs may be regulated mainly by posttranslational protein modifications (such as phosphorylation/ ubiquitination), or by protein-protein interactions, in which the ORs activated/inactivated by interacting with other co-activator/co-suppressor factors (Giguere, 1999).
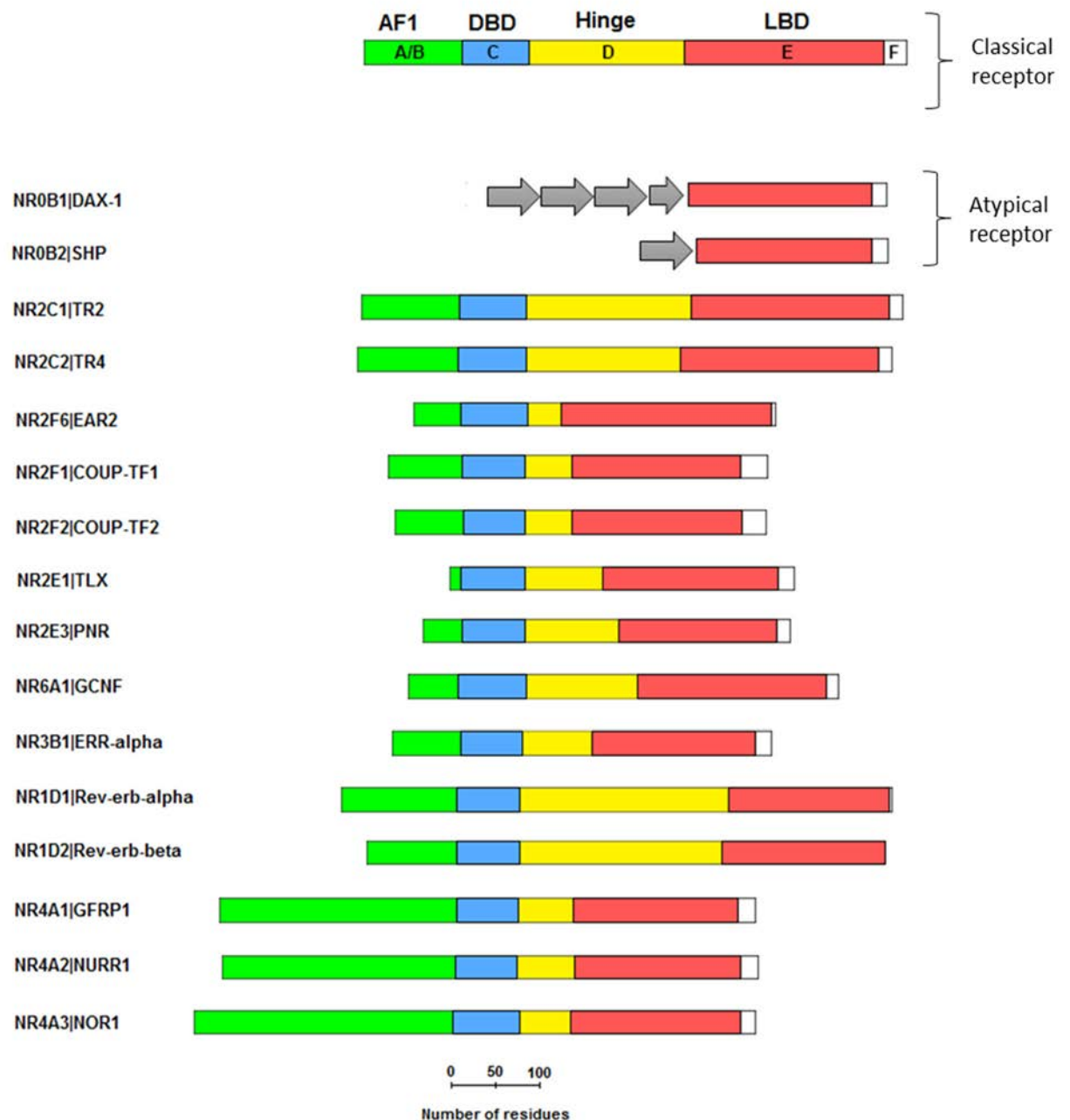
Figure 1.11 The protein structures of human orphan receptors. The classic structure of ORs are composed of six domains: domain A/B contains an activation domain (AF1; green box) of variable length, domain C is the DBD (blue box), domain D is a hinge domain (yellow box), domain E (red box) contains the dimerization interface, LBD, and domain F is a variable region of unknown function (white box). In two atypical nuclear receptors belonging to the NR0B family, the A/B, C, and D domains are replaced by a 65 to 70 amino acid motif rich in alanine and glycine (grey arrow). This motif is repeated 3.5 times in NR0B1 and only once in NR0B2.

The DBDs of ORs interact with specific DNA sequences of the regulatory regions of target genes, known as Hormone Response Elements (HREs). These elements consist of one or two half-site motifs (Claessens and Gewirth, 2004). In general, ORs bind to the direct repeats elements (DRs) as homodimers, where their common half-site motif sequence on direct repeats is 5'-AGGTCA-3' (Mangelsdorf *et al.*, 1995; Sandelin and Wasserman, 2005). However, some ORs can make heterodimers with RXRs such as NR4A1, and NR4A2 (Perlmann and Jansson, 1995). The NR2F/C classes of orphan nuclear receptors show similar DNA binding sites. The various ways that ORs bind to DNA are shown in Figure 1.12.

NR2C1/2 (TR2 & TR4) proteins bind to DNA as homodimers, but they only recognise the AGGTCA n AGGTCA as the direct repeat elements (DRs), 'n' is a nucleotide spacer that can vary between 0-6 nucleotides (Yan *et al.*, 1998; Benoit *et al.*, 2006). The NR2F1/2 (COUP-TFs) proteins also bind to DNA as homodimers but they recognise RGGTCA n RGGTCA as the direct repeat elements (DRs), but they also can interact with the inverted repeat elements (IRs) RGGTCA n TGACCY, 'n' ranges between 0 and 11 nucleotides, where R=A/G, and Y=T/C (Cooney *et al.*, 1992; Mangelsdorf *et al.*, 1995; Benoit *et al.*, 2006).

Interaction of ORs with DNA is restricted by their dimerization modes as whether they bind as homodimers or heterodimers, but also because some ORs exhibit the ability to bind one half-site motif as monomers, such as NR4A subfamily (Wilson *et al.*, 1993). NR4A subfamily proteins bind to half-site sequence AAAGGGTCA as monomers, which is usually proceeded with an A/T reach sequence (Wilson *et al.*, 1991). NR1D1/2 bind to DNA as homodimers or monomers, to more flexible half-site sequence WAWNTRGGTCA, where W= A/T, and R=A/G (Harding and Lazar, 1993).
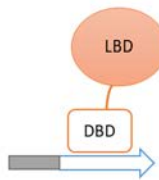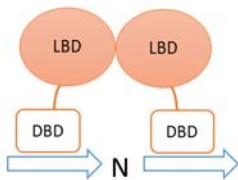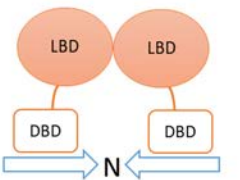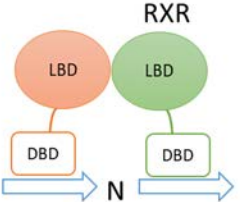
Figure 1.12 DNA binding moders of Orphan Nuclear Receptors. Orphan receptors can bind to hormone response elements as monomers, homodimers, and RXR-heterodimers.

1.5.2 The role of orphan nuclear receptors in the ALT mechanism

Studies have shown that the presence of TCAGGG variant repeats (C-type repeats) within telomeric DNA of in a WI38VA13_2RA (an ALT+ cell line) recruits the NR2C/F classes of ORs, particularly NR2F2 and NR2C2. It has also been suggested that the association of NR2F2 and NR2C2 proteins with ALT+ telomeres may alter the chromatin by recruitment of the NuRD nucleosome remodelling and histone deacetylation complex and by recruitment of a zinc-finger protein ZNF827 in cell-cycle-dependent manner (Conomos *et al.*, 2012; Conomos *et al.*, 2014; Marzec *et al.*, 2015). In WI38VA13_2RA cells, ALT markers were significantly correlated with localisation of NR2C/F nuclear receptors at telomeres. The depletion of NR2F2 in these cells resulted in significant reduction in number of APBs, telomere shortening, and an increase in the expression of NR2C2, which leads to the suggestion that they play role in ALT activity (Dejardin and Kingston, 2009; Conomos *et al.*, 2012). It was proposed that the recruitment of NR2F and NR2C classes of ORs to telomeres disrupts the binding of shelterin, and changes the telomeric chromatin status in a way that facilitates interaction between telomeres, thereby promoting strand invasion and ALT activity in APBs (Conomos *et al.*, 2012; Conomos *et al.*, 2014; Marzec *et al.*, 2015) as shown in Figure 1.13. However, it is not known if this relationship with these ORs is a general feature of all ALT+ cells and tumours and therefore a requirement for ALT.
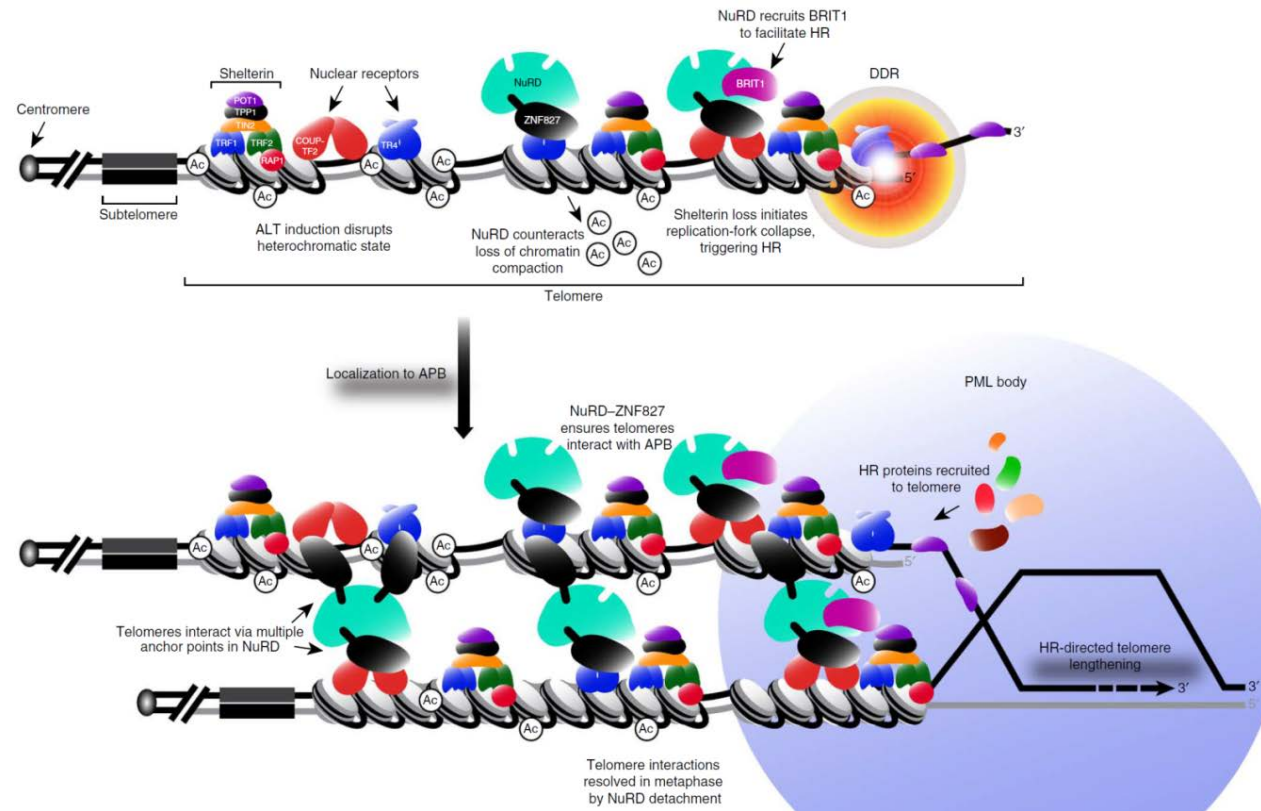
Figure 1.13 Model for telomere-telomere interaction and telomere lengthening orchestrated by association with the NR2C/F OR subfamily of proteins at telomeres in ALT cells (Conomos *et al.*, 2014). It proposes that localisation of NR2F2/COUP-TF2 and NR2C2/TR4 at telomeres of WI38VA13_2RA cells facilitate the physical intersaction between telomereses throughtout the NuRD-ZNF827 complex , thereby promoting strand invasion and ALT activity in APBs.

1.5.3 Werner Syndrome derived ALT+ cell lines and ORs

Werner syndrome is a rare autosomal recessive disorder associated with a premature aging phenotype. Germline mutations in *WRN* cause Werner syndrome. *WRN* is ∼ 140 kb in length, it lies on human chromosome 8 and encodes the WRN protein (Goto *et al.*, 1992). WRN has been shown to play roles in resolving secondary structures of telomeres and it interacts with TRF2 so promoting the disassembling of t-loops in cell-cycle-dependent manner (Opresko *et al.*, 2004).

The WV cell line was derived from person with Werner syndrome ($WRN^{-/-}$), it was SV40-transformed and subsequently became immortalised as an ALT + cell line. Analysis of XpYp telomere mutations in this cell line showed loss of variant or degenerate repeats compared to the progenitor allele and replacement with canonical (TTAGGG)$_n$ repeats (Figure 1.14) (Mendez-Bermudez *et al.*, 2012). This suggests that potential binding site for ORs were being lost from XpYp telomere molecules in the WV cell line.

Figure 1.14 XpYp telomere mutations in the WV cell line. This figure shows mutations analysis at one haplotype of XpYp telomeres in the WV cell line. The distribution of TTAGGG repeats represented by (T), while variant repeat sequences TGAGGG resembled by (G) (Mendez-Bermudez *et al.*, 2012).

1.6 Functional analysis of genes

1.6.1 Functional analysis of genes using RNA interference

The RNA interference (RNAi) is an endogenous biological pathway that employs noncoding microRNAs (miRNAs) to regulate gene expression by neutralizing targeted mRNA (Figure 1.15). In mammalian cells, each miRNA is transcribed as a hairpin precursor (primary-miRNA), which is processed by an RNAase-III enzyme called Drosha into ~70 nucleotides (pre-miRNA). These pre-RNAs are subsequently transported to the nucleus to be cleaved into short double-stranded fragments of ~21 nucleotides (siRNAs) by an RNAase-III-like enzyme called Dicer (Hammond *et al.*, 2000; Elbashir, Harborth *et al.*, 2001; Elbashir, Lendeckel *et al.*, 2001; Lee *et al.*, 2003). In most cases, the mature siRNA is unwound into two single stranded RNAs, the anti-sense (guided) strand and the sense (passenger) strand. The guide strand is incorporated into the multiprotein RNA-induced silencing complex (RISC), while the passenger strand is quickly degraded. The RISC complex mediates the annealing of the anti-sense strand to complementary sequences of the target mRNA facilitating its translation attenuation or degradation (Hammond *et al.*, 2000; Birmingham *et al.*, 2007). The catalytic component of RISC, Argonaute 2 (Ago2) is required for the cleavage of the target mRNA in human and other organisms. This form of post-transcriptional gene regulation can be used to investigate gene function as synthetic short duplexes of 21-25 nt siRNAs can be incorporated into the RISC directly and alter transcript levels by rapid degradation of target mRNA.
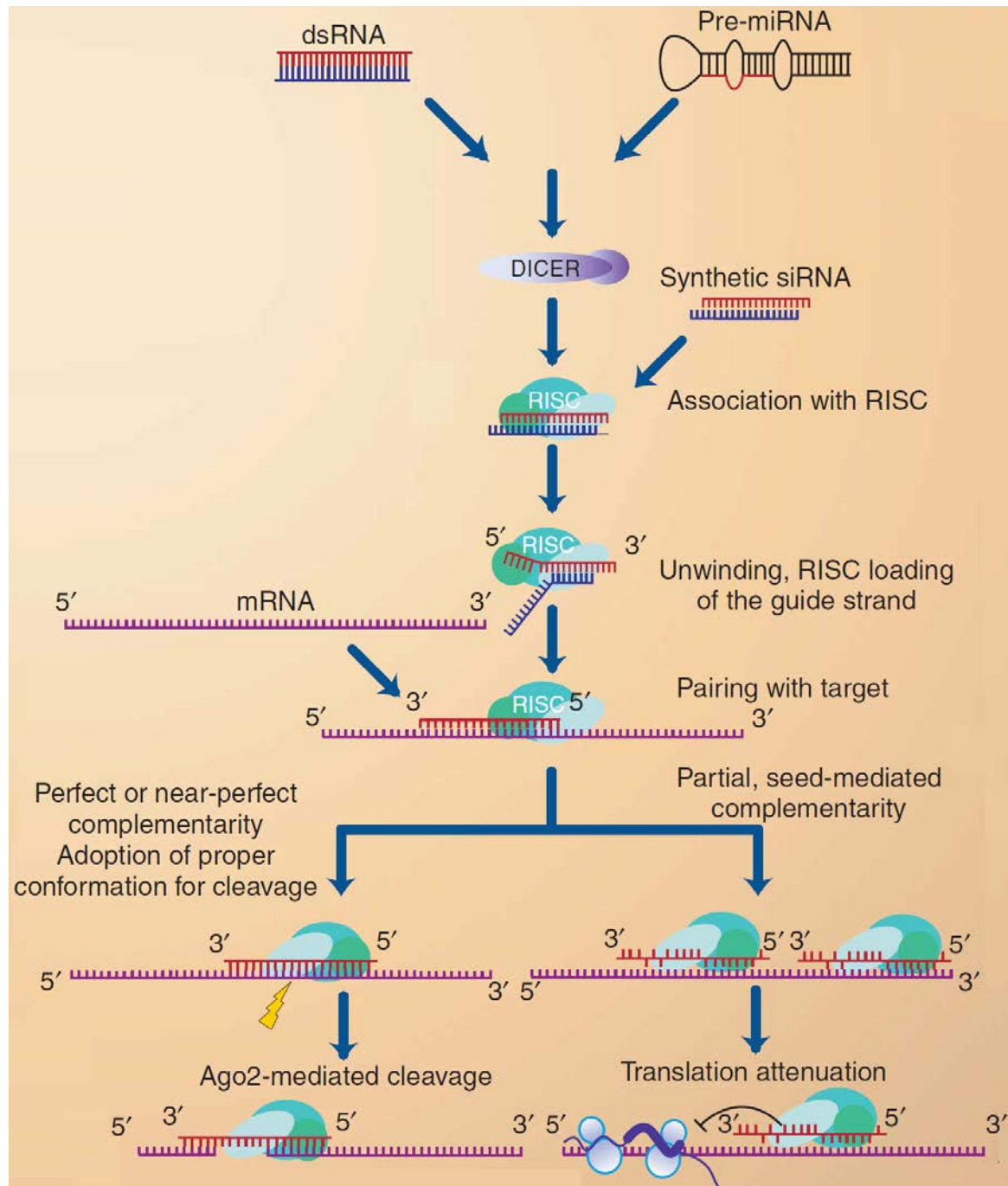
Figure 1.15 Schematic representation of mammalian RNAi pathway. The roles of Dicer and RISC in processing a range of different substrates for the RNAi mechanism are shown. Adopted from (Birmingham *et al.*, 2007).

1.6.2 Functional analysis of genes using CRISPR-Cas9 editing system

1.6.2.1 Introduction to CRISPR

The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) as a term was firstly named by Jansen *et al.* (2002). However, these interspaced short palindromic repeats was first discovered by Nakata et al. (1989) in the *E. coli* genome. Since then, similar loci were found in the genomes of most prokaryotes (Bacteria and Archaea), but not in eukaryotes or viruses (Mojica *et al.*, 2000). In each species, the consensus length of these repeats ranged between 21-40 bp, and are interspersed with similarly-sized, non-repetitive sequences (spacers). These loci are flanked one side of cluster by a 300-500bp region termed as leader sequence (Jansen *et al.*, 2002). The direct repeats and the flanking leader sequences were different between species, but conserved within a species. The possibility of existing multiple CRISPR loci per genome suggested that the CRISPR locus is a mobile element. Jansen *et al.* (2002) also identified four genes that were associated with the CRISPR loci in the prokaryotes. These genes were named as Cas (CRISPR-Associated Sequence) genes. Furthermore, three types (I-III) of CRISPR-Cas were identified in prokaryotic genomes (Haft *et al.*, 2005).

After that, it was found that the non-repetitive spacers are exogenously derived sequences (Pourcel *et al.*, 2005), suggesting that CRISPR-Cas is a microbial immune system against invading viruses (Barrangou *et al.*, 2007). The transcription of the CRISPR locus into non-coding RNA (crRNA) provides an adaptive immunity to viruses and plasmids when the spacer sequence is homologous to the exogenous target sequence named as protospacer. The protospacer sequence can be DNA sequence (Marraffini and Sontheimer, 2008) or an RNA sequence (Hale *et al.*, 2009). It was shown that each protospacer is associated with a protospacer adjacent motif (PAM), which is required for the recognition of the protospacer and stimulating the cleavage by Cas (Garneau *et al.*, 2010). The understanding of the CRISPR-Cas system was further developed by Deltcheva *et al.* (2011). Their study for type II CRISPR-Cas system (the most frequently used) in *Streptococcus pyogenes* identified the role of trans-activating CRISPR RNA (tracrRNA) in processing the pre-crRNA into discrete units. Furthermore, it was found that the mature crRNA that is base-paired to tracrRNA forms a two-RNA heteroduplex (tracrRNA:crRNA) that directs Cas nuclease activity to cleaved the target DNA in the

presence of PAM (Jinek *et al.*, 2012). It also showed that the tracrRNA:crRNA can be substituted with one synthetic chimeric RNA (termed "single guide RNA" or gRNA). The gRNA is composed of a 20 nt region homologous to the target DNA and a scaffold region that provides structure. The PAM in the target DNA sequence can vary depending on the specific CRISPR system. The SpCas9 protein (derived from *S. pyogenes*) recognises a 5′-NGG-3' PAM in target DNA.

1.6.2.2 Using type II of CRISPR-Cas9 system in editing of mammalian cells

The genome editing using type II of CRISPR-Cas (CRISPR-Cas9) system as a programmable endonuclease is one most powerful technology that can be used to modify the genome in a pre-designed manner. To use CRISPR-Cas9 system in eukaryotic cells, both non-coding gRNA and mRNA of Cas9 need to be expressed, where the Cas9 nuclease activity guided by gRNA inducing cleavage of the genomic DNA (Figure1.16).
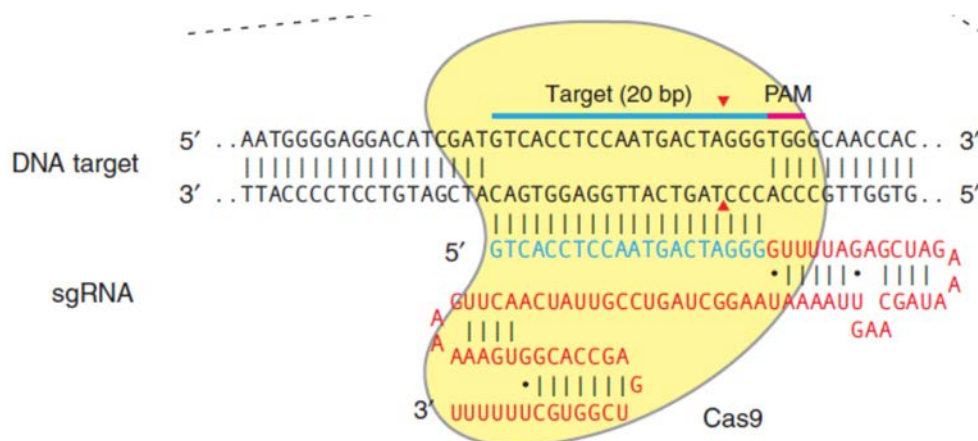
Figure 1.16 Schematic of genome editing using CRISPR-Cas9 System with a single guide RNA.The synthetic guide RNA (sgRNA) is composed of a 20 nt guide sequence (blue) and a scaffold (red). The guide sequence part of sgRNA base-pairs with the DNA target (blue bar) directly of 5′-NGG-3' protospacer adjacent motif (PAM). The wild type Cas9 protein from *S. pyogenes* (yellow) recognises this interaction (gRNA:DNA heteroduplex) and cleaves 3 bp upstream of the PAM (red arrows). Adopted from Ran *et al.* (2013).

The CRISPR-Cas9 as other genome editing technologies utilises a cellular DNA repair pathway in order to achieve a desired editing outcome. Following cleavage by

Cas9, the target locus is ideally processed by one of two major DNA repair pathways; the NHEJ or the HR pathway. NHEJ is known as an error-prone DNA repair pathway that can introduce variable length of insertion/deletion (indel) mutations at the site of DSBs. NHEJ is used by cells to repair the DSB in in the absence of a homologous template in any phase of cell cycle except S phase (Lieber, 2010). Exogenous HDR templates can be designed and introduced along with Cas9 and sgRNA to promote precise editing sequence alteration at a target locus. In the absence of a repair template, NHEJ can be used for gene disruption by introducing indel mutations in a protein-coding exon that lead to frameshift and a pre-mature stop codon (Figure 1.17).

Targeting early exons with CRISPR-Cas9 mediated frameshift indel mutations is known to be associated with triggering nonsense-mediated decay (NMD), which is equivalent to a gene deletion as the mRNA would never be translated (Popp and Maquat, 2016). The understanding of NMD can be used for better editing efficiency of gene knockout.
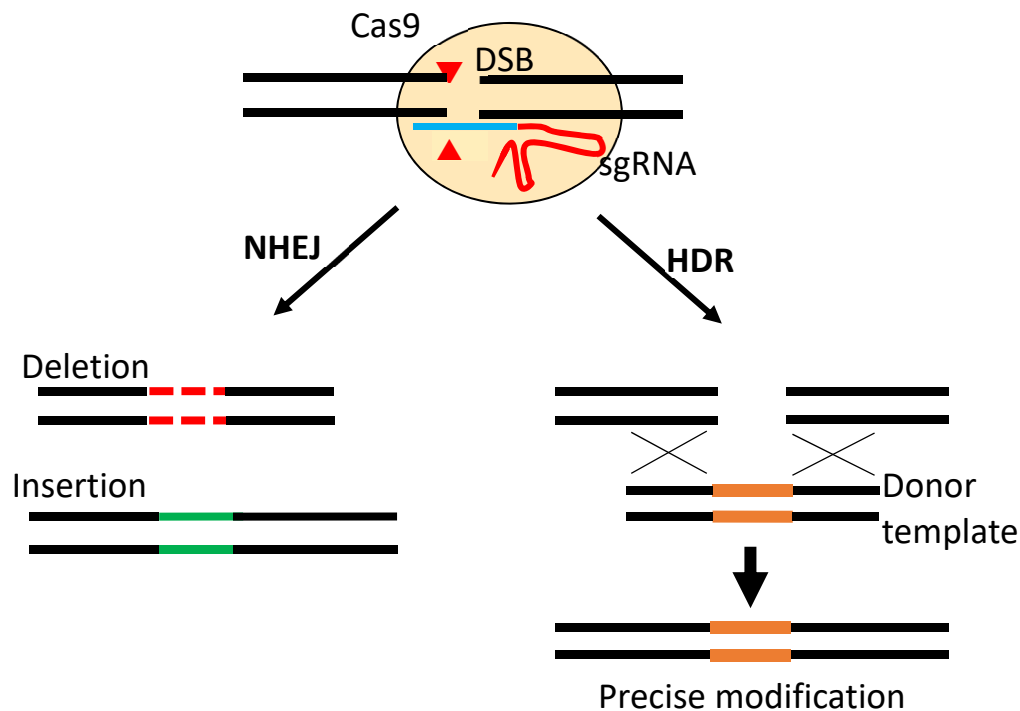
Figure 1.17 Genome editing using CRISPR-Cas9 system is mediated by repair of nuclease-induced DSB by NHEJ or HDR pathways.NHEJ repair of DSB can introduce variable length insertion or deletion (indel) mutations. The HDR repair of DSB can produce precise modification using a donor DNA template. Note: sgRNA, synthetic single guide RNA; DSB, double-strand break; HDR, homology-directed repair; NHEJ, nonhomologous end joining. Adopted from Sander and Joung, (2014).

1.6.2.3 Off-targets of CRISPR-Cas9 system

The Cas9, as other endonucleases cleaves off-target sequences at lower frequencies. The positions of tolerated base mismatch vary in a 20 nt guide RNA sequence, but more likely to occur distal of PAM (5' of the PAM), while perfect base-pairing within 10 to 12 bp directly upstream 5' of the PAM showed to determines the specificity of Cas9. Overall, Cas9 can tolerate up to 5 nt mismatches in 20nt guide RNA sequence (Fu *et al*. 2013; Hsu *et al*. 2013; Fu *et al*. 2014).

Modifications of Cas9 has showed promising solutions for the Cas9 off-targets problem. Inactivating mutation in one of the Cas9 domains showed to results in Cas9 with nickase activity named Cas9 D10. In contrast to DSB, the individual nicks in the genome can be repaired with high fidelity. The double nickase Cas9 system has successfully used for genome editing with reduced the off-target effects by > 50 fold in cell lines (Ran, Hsu, Lin *et al.*, 2013). Recently, a high fidelity version of Cas9 (SpCas9-H1) has been developed that harbouring modifications that increase the specificity of recognition of the 20nt target sequence in the genome. The off-target assessment by genome-wide screening for the potential off-targets following editing by SpCas9-H1 shows nearly no off-target effects detectable in non-repetitive regions (Kleinstiver *et al.*, 2016).

1.7 Aims

ALT is a homologous recombination-based mechanism with a poorly understood molecular basis but it is used by some cancers, in particular those with a mesenchymal origin such as sarcomas. The first aim of my research project was to investigate the expression of selected orphan nuclear receptors, their telomere localisation and the abundance of potential binding sites for these proteins within telomeric DNA in a panel of ALT+ cell lines. Following this, the regulatory role NR2F2 was investigated by transient down regulation of NR2F2 protein expression in ALT+ cell lines, including a Werner syndrome derived ALT+ cell line, by measuring changes in phenotypic markers of ALT, cell cycle analysis, and gene expression.

The second aim of my research project was to utilise the CRISPR/Cas9 genome editing system to mutate *ATRX* in a pre-crisis SV40-transfomred fibroblast clone. Subsequently the aim was to use a wide variety of methods to characterize the telomere maintenance mechanism in populations of cells that survived telomere-driven crisis. Followed by integrated analyses of whole-exome sequencing (WES) and transcriptome profiling (RNA-seq) to uncover genetic and gene expression changes associated with activation of the ALT mechanism.

Chapter 2: Material and methods

2.1 Materials

2.1.1 Oligonucleotides

PCR primers (18-24 nt) and CRISPR gRNA cloning (41-42 nt) oligonucleotides were synthesised by Sigma-Aldrich Company (Poole, UK), while stealth siRNA (25 nt) oligonucleotides were synthesised and pre-annealed by Life Technologies. See Appendix 2.1.

2.1.2 Antibodies

The following primary antibodies were used in this work: NR2C2/TR4 and NR2F2/COUP-TF2 (from Perseus Proteomics, Japan), ATRX (H-300) and PML (from Santa Cruz, USA), GAPDH (from Ambion, UK). The following four secondary antibodies were used: two horseradish peroxidase-conjugated antibodies (GE Healthcare, UK), and two fluorophore-conjugated Alexa488 antibodies (from Abcam, UK). See Appendix 2.2 for more details.

2.1.3 Antibiotics

The following antibiotics were used in this work: Puromycin dihydrocloride, G-418 disulfate salt, and Ampicillin (all obtained from Sigma-Aldrich Company, UK).

2.1.4 Cell lines

Table 2.1 gives information about the primary and established cell lines used in this research project.

Table 2.1 Cell lines used in this work.

Cell line	Sex	Tissue Origin	Immortalisation status	Telomere maintenance	Source
HT1080	Male	Fibrosarcoma	Tumour	Tel+	European Collection of Cell Cultures (ECACC)
WV	Male	Fibroblast, Werner syndrome	SV40	ALT+	Derived from a normal fibroblast cell line from a 45-year-old male patient with Werner's syndrome that was transformed by SV40 infection. A gift from Roger R. Reddel
WI38VA13_2RA	Female	Fibroblast, lung	SV40	ALT+	European Collection of Cell Cultures (ECACC)
SUSM-1	Male	Fibroblast, liver	Chemical	ALT+	A gift from Professor Olivia M. Pereira-Smith
SAOS-2	Female	Osteosarcoma	Tumour	ALT+	A gift from Professor Paolo Salomoni
U2OS	Female	Osteosarcoma	Tumour	ALT+	A gift from Professor Paolo Salomoni
CCD18LU/TCL1	Female	Fibroblast, Lung	Primary cell line – with extended lifespan by SV40 Large T; not immortal.	none	Generated from the primary lung fibroblast cell line CCD18Lu (ECACC); transformed by transfection with SV40 Large T antigen and clone TCL1 isolated (Royle Laboratory)

2.2 Methods

2.2.1 Cell Culture

2.2.1.1 Culturing cell lines

Three different types of media were used for culturing six established human cell lines. HT1080, W-V, and U2OS cell lines were grown in DMEM medium (Gibco, UK) with 10% Fetal Bovine Serum (Gibco, UK). The SAOS2 cell line was grown in RPMI 1640 (Gibco, UK) supplemented with 10% FBS, while WI38VA13_2RA was grown in MEM medium (Gibco, UK) supplemented with 10% FBS and 1X Non-Essential Amino Acids (1XNEAA; Gibco, UK).

Cells were subcultured 1:2, 1:3 or 1:4 when at 80-100% confluency, which depended on the growth rate of each cell line and the purpose of the seeding. The subculturing was achieved by removing the media from the T-25 flask and briefly washing the adherent cells by adding 1 ml of pre-warmed 1X Trypsin-EDTA. The cells were then detached by incubating them at 37°C for 2-3 minutes in another 1 ml of pre-warmed 1X Trypsin-EDTA followed by the addition of 7 ml of pre-warmed medium to inactivate the trypsin. Cells were grown in a 37°C incubator with ambient O_2 (21%), CO_2 (5%). Similar procedure was used for culturing CCD18LU primary cells with DMEM medium containing 20% Fetal Bovine Serum (FBS) and 400 µg/ml of G418 to maintain expression of SV40 large T antigen. Primary cell lines were grown at 37°C in normoxia conditions, 5% O_2 and 5% CO_2.

To prepare cell pellets for downstream analyses, trypsinised cells in medium were transferred to falcon tubes, and centrifuged of 1000 rpm at 20°C for 8 minutes. The supernatant was aspirated and the cells were washed and re-suspended with the appropriated amount of 1X sterile PBS. A small volume of cell suspension was used for cell counting. The tubes of cells were centrifuged again, the supernatant was aspirated and cell pellets were transferred into dry ice, until placed in a -80°C freezer.

2.2.1.2 Cell counting and Population Doubling Calculation

Cell counts were performed using a haemocytometer and mean population doubling (PD) was calculated using the equation:

$$PD = \frac{[\ \log 2\ (\text{count at harvest}) - \log 2\ (\text{initial cell number})\ ]}{0.301}$$

2.2.1.3 Cell viability assay

Trypan blue is a dye that selectively stains dead cells, while uptake is excluded from living cells, allowing an estimation of the percentage of viable cells. Cell suspensions were incubated with an equal volume of Trypan Blue buffered solution (Gibco, UK) (FC 0.2%) for 3 minutes and counting was carried out using a Haemocytometer. Cell viability was calculated using the following equation:

$$\% \text{ Viable cells} = \frac{\text{Number of dead cells (blue stained)}}{\text{Total number counted of Cells}} *100$$

2.2.1.4 Cloning of CRISPR-Cas9 vectors

Plasmids pLEICS-97, pLEICS-122 were used to express the wild type SpCas9, non-coding RNA of selected CRISPR target sites in ATRX-Exon 9 (CRISPR-TS1, CRISPR-TS2), and the puromycin resistance gene for antibiotic selection purposes. The cloning of selected single guide RNA sequences was carried out by the Protein Expression Laboratory (PROTEX) facility, University of Leicester. PROTEX used a recombination-based cloning method for inserting the U6 promotor-sgRNA sequence into the family 95 vectors (based on homology sequence). This approach is based on site-specific recombination (Walhout *et al.*, 2000). In this cloning approach, the Forward oligo (42-nt, containing the reverse complementary of 20nt gRNA sequence) and the Reverse primer (41-nt, containing the 20-nt guide sequence) were designed as advised by PROTEX and synthesised (see Appendix 2.3). A PCR using the forward and reverse oligonucleotides was conducted by PROTEX to generate two inserts that overlapped in sgRNA sequence and shared site-specific homologous regions in Family 95 vectors. The final plasmid constructs (9183 bp), following recombination-based cloning, showed replacement of the T7 promotor and SacB gene with the U6-sgRNA sequence.

Plasmids were received as solid and liquid media cultures of *E. coli* (provided by PROTEX for Family 95-after cloning, or by Addgene for PX495v2.0 [Empty CRISPR-Cas9 vector]). Bacteria were streak on agar plates contained 100 μg/ml ampicillin and incubated upside down at 37°C overnight (~ 16 hours) in an aerobic conditions. Single colonies were grown in 5 ml of Luria Broth (LB) contained 100 μg/ml ampicillin at 37°C in a shaking incubator for ~16 hours. The culture (4.5 ml) was used for isolation of DNA via the plasmid DNA miniprep kit (OMEGA), while the

500μl was added to equal volume of 50% glycerol solution for long-term storage in -80 °C. The plasmids were verified by restriction digestion, and Sanger sequencing of PCR amplicons generated around (1) the 20nt guide sequence, (2) the SpCas9, and (3) Puromycin resistance gene.

2.2.1.5 Delivery of CRISPR/Cas9 vectors to SV40-transformed pre-crisis fibroblast cells

A large quantity of plasmid DNA was isolated from verified bacterial colonies using the plasmid DNA maxi prep kit (OMEGA) following the manufacturer's instructions. Plasmid DNA concentration was estimated using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, UK) and plasmid quality was assessed by agarose-gel electrophoresis. Plasmid DNA was concentrated by ethanol-precipitation and resuspended in sterile H_2O (Sigma-Aldrich, UK).

The delivery of CRISPR/Cas9 plasmids to the CCD18LU/TCL1 cells was carried out using an Amaxa Nucleofector (an efficient electroporation delivery system). Prior to transfection, cells were subcultured in T-75 flaks in order (1) to maximise the proportion of cells in their logarithmic growth phase, (2) to collect conditioned media and (3) to generate the required number of cells for parallel transfection reactions. On the day of transfection, cells were trypsinised and aliquots of $1X10^6$ cells were pelleted at low-speed centrifugation (90 x g) for 10 min. Each cell pellet was re-suspended with 100 μl of Ingenio transfection solution (Mirus, UK), and 2 μg of linearised plasmid DNA was added to each transfection reaction. Electroporations were carried out in 0.2 cm cuvettes using an Amaxa program and as recommended by the manufacturer. For optimising the delivery efficiency five Amaxa programs (A-24, U-12, U-23, T-16, and V-13) for electroporation of mammalian primary cell lines were assessed for delivery efficiency by transfecting cells with a red fluorescent protein (RFP) expression plasmid (RFP1-pcDNA3.1/zeo , ~6 kb). Cell viability was assessed by counting cells with Trypan blue staining. The delivery frequency measured by transfection with the RFP plasmid was calculated:

$$\% \text{ Delivery Efficiency} = \frac{\text{Number of nuclei with RFP signal}}{\text{Total number of nuclei counted}} *100$$

2.2.1.6 Cloning of primary cells

After 48 hours of transfection, media were removed and fresh media with Puromycin selection (0.7 µg/ml) were added. Media containing Puromycin was changed twice in the first week, and selection was removed from the media in the second week. Two weeks later, selected single clones were expanded into 24-well plates using Scienceware® cloning discs (Sigma-Aldrich). Clones that continued dividing were transferred to 6 well-plates, and cells were subsequently expanded into T25 flasks. During subculturing, cell pellets were collected from the clones that continued to divide.

2.2.1.7 Delivery of siRNA using lipid system

Lipofectamine® RNAiMAX Reagent with Opti-MEM® Medium (Invitrogen) was used to deliver the stealth siRNA into human cell lines for transient downregulation of gene expression. Cells were grown in 10cm petri-dishes (each contained 3 small coverslips) with fully supplemented medium for 24 hours to reach approximately 40-50 % confluence before the delivery of siRNA. Transfection was carried out following the manufacturer's instructions. Prior to transfection, the medium was replaced and different concentration of dsRNA oligonucleotides (25 to 50 pmol) were mixed with appropriate volumes of low toxicity Lipofectamine® RNAiMAX Reagent in Opti-MEM® Medium. The mixture was added to growth medium in a 1:5 ratio. After 72 hours, the coverslips were removed for APBs analysis. The rest of the cells were detached from the petri-dish by trypsinisation and harvested for different assays: ~$1X10^6$ cells for C-circle assays, $0.5x10^6$ for western blot analysis, $1.5x10^6$ RNA-Seq, $1X10^6$ for cell cycle analysis by fluorescence-activated cell sorting (FACS) analysis.

2.2.1.8 FACS analysis of the cell cycle

Cell cycle analysis of un-synchronised cells was carried out with Propidium Iodide (PI) DNA staining and FACS. Approximately 1×10^6 cells were harvested after transfection and pelleted by centrifugation at 2,000 rpm for 5 min at 4°C. The cells were fixed and permeabilised by re-suspending in 1ml of 70% ice-cold ethanol and stored at -20°C until analysed. On the day of analysis, cells were washed with 1X PBS and centrifuged at 2000rpm for 5 minutes at 4°C, a re-suspended in 491µl of 1X PBS and transferred to a FACS tube. DNA was stained with 4µl PI buffer solution (stock 2.5mg/ml), and RNA was degraded by adding 5µl RNAase (stock 10mg/ml). Samples were incubated at 37 °C for 30 min. Cells were processed for FACS measurements

conducted using BD FACSCanto™ II, and the analysis of cell cycle distribution was carried out using the BD FACSDiva Software. Microsoft Excel and R v3.4.1 programs were used for statistical analysis. ANOVA and pairwise multiple comparison T-tests (with Benjamini-Hochberg correction) were applied to determine significant differences when the P value < 0.05.

2.2.1.9 Annexin V/Propidium Iodide apoptosis assay

Cells that activated the caspase system are characterised by expressing phospholipid phosphatidylserine to the external cellular environment that can be detected by FITC-conjugated -Annexin V which is Ca2+-dependent phospholipid-binding protein with high binding affinity for phospholipid phosphatidylserine. For differentiation between the early and late apoptosis the FITC-conjugated-Annexin V used in combination with the propidium iodide (PI) dye that selectively uptake by cells that loss of the membrane integrity indicating late crisis. The Annexin V Apoptosis Detection FITC Kit (eBioscience, UK) was used for this assay and following the manufacturer's instructions. Cells were processed for FACS measurements conducted using BD FACSCanto™ II, and two staining parameters (PE, FITC) settings were applied using the BD FACSDiva Software.

2.2.2 Molecular biology

2.2.2.1 Western blotting

2.2.2.1.1 SDS-polyacrylamide gel electrophoresis

Cell pellets of ~$1X10^6$ cells were lysed with an appropriate volume of whole cell lysis buffer (RIPA buffer) by vortexing and incubation for 30 minutes on ice. The lysates were clarified by spinning at 12,000 rpm for 15 min, at 4°C. The Bradford assay was used to determine protein concentrations, and proteins were diluted to 1-3mg/ml in Laemmli loading buffer (FC: 63 mM Tris HCl, 10% Glycerol, 2% SDS, 0.0025% Bromophenol Blue, 2.5% β-mercaptoethanol). Electrophoresis of proteins was carried out in SDS-polyacrylamide gels. Prepared lysates were boiled at 96°C for 5 min to unfolding proteins prior to loading onto a polyacrylamide gel. The proteins were migrated through a stacking gel of 3% acrylamide, then size separated using a 6 -12%

acrylamide gel, which varied between the experiments depending on the size of the protein of interest. Electrophoresis was accomplished at 100 volts for approximately 1 hour 30 minutes.

2.2.2.1.2 Transfer of proteins onto a nitrocellulose membrane

The protein transfer was carried out in semi-dry conditions (except ATRX, which was conducted in wet-transfer conditions). The gel and a nitrocellulose membrane (Amersham Hybond ECL, GE Healthcare) were soaked for 10 minutes in 1X transfer buffer (25mM Tris, 192mM glycine, 20% methanol), or in the case of ATRX the membrane was pre-soaked in transfer buffer excluding methanol and with the addition of 0.01% SDS. The protein transfer was carried out in a semi-dry transfer unit (BioRad) at $1mAmp/cm^2$ for 50 minutes. Blots were stained with Ponceau staining solution (Sigma) for 5 minutes with agitation to verify protein transfer prior to immunoblotting.

2.2.2.1.3 Immunoblotting

Blocking buffer was made fresh every time by dissolving 5% w/v skimmed milk in 1X PBST (1X PBS contained 0.1% Tween 20) and added to the western blots for a minimum of 1 hour. The primary antibody was diluted in blocking buffer (see Appendix 2.2 for dilutions used) and added to blots for a 1 hour incubation at room temperature or overnight at 4°C with agitation. This was followed by three washes with 1X PBST to remove the excessive primary antibody. The secondary antibody (1:10,000 horseradish peroxidase-conjugated antibody in blocking buffer) was applied for 1hr, followed by three washes with 1X PBST. Blots were then processed for protein detection using the luminol substrate in the ECL (enhance chemiluminescence) solution (Amersham ECL Prime Reagent, GE Healthcare) and following the manufacturer's instructions. After that, blots were exposed to X-ray film (Kodak) and developed.

2.2.2.1.4 Quantitative analysis of western blots

For protein quantitative analysis, X-ray films were scanned to generate TIFF image files with 300 pixels resolution. The area measuring of each track was collected using ImageJ v1.5e software. The data normalisation was carried out as described in (Degasperi *et al.*, 2014). Microsoft Excel and R v3.4.1 programs were used for statistical analysis and generating histograms of the expression level relative to the control protein. ANOVA and pairwise multiple comparison T-tests (with Benjamini-

Hochberg correction) were applied to determine significant differences when the P value < 0.05.

2.2.2.2 Immunofluorescence microscopy

2.2.2.2.1 Immunofluorescence and Fluorescence *In Situ* Hybridization Protocol

Cells were grown on 1.3cm coverslips until they reached 70-90% confluency Coverslips were washed briefly in 1XPBS and then cells were fixed for 15 minutes in 3.7% formaldehyde in PBS. Coverslips were then washed three times in PBS to remove the excess of formaldehyde. Cells were permeabilised in 0.5% Triton X100 in 1XPBS for 5 minutes, and then incubated for 1 hour in 3% BSA in PBS. Cells were incubated with primary antibody diluted in 3% BSA in PBS (see Appendix 2.2 for used dilutions) for a minimum of 1 hour at room temperature or overnight at 4°C. Coverslips were washed three times in PBS solution containing 0.1% Triton to remove excess primary antibody prior to a 30 minutes incubation with the AlexaFluor488-dye conjugated secondary antibody (1:1000 dilution in blocking buffer) for 1 hour at room temperature. The washes were repeated as before, but the coverslips were protected from light. Processing of the coverslips ended by mounting on microscope slides with ProLong Gold Antifade Mountant (Life Technologies) that contained 5µg/ml of DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride), or processing continued in order to combined the IF detection protocol with Fluorescence *In Situ* Hybridization (FISH) for telomere colocalisation analysis using the following steps. Cells were fixed again with 4% formaldehyde in 1X PBS for 10 minutes, and then washed twice with 1 x PBS. Cells were then dehydrated by washing the coverslips in increasing concentration of ethanol (70%, 80% then 100%), and then allowed to air dry. Cells were incubated with 50nM Cy3–OO-(CCCTAA)$_3$ PNA probe (Panagene) in the PNA hybridization solution (70% deionized formamide, 0.25% blocking agent (PerkinElmer), 10mM Tris-HCl, pH7.2). The edges of the coverslips were sealed with vulcanising rubber solution. DNA denaturation was carried out using Hybaid OmniSlide Thermal Cycler (Thermo Scientific) and cycled as follows: 80°C for 3 minutes, 22°C for 2 hours. The excess PNA probe was removed by washing in formamide-Tris solution (70% formamide, 10 mM Tris-HCl pH 7.2), then in a Tris-Tween solution (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.05% Tween). Finally, coverslips were washed briefly with 1XPBS, dehydrated by 1 minute incubations in 70%, 80% then 100% of ethanol, and left to air dry.

Coverslips were mounted on microscope slides with ProLong Gold Antifade Mountant contained 5µg/ml of DAPI.

2.2.2.2.2 Microscopy and image processing

Samples were visualised for colocalisation using an Olympus FV1000 confocal microscope. Nuclei at interphase were captured in 15-20 Z planes in 0.25µm increments for the green (AlexaFluor488), red (Cy3- Tel-PNA) and blue (DAPI) channels. For quantitative analysis, slides were screened using the Olympus Cell^R/scan^R system (Advanced Imaging Facility, University of Leicester) to capture images of 100s-1000s of nuclei. Approximately 100 interphase nuclei were captured per coverslip in blue, green and red channels, and analysis of co-localisation events was conducted on the captured images using ImageJ v1.5e software with colocalisation java plugin tool written by Pierre Bourdoncle, Website "https://imagej.nih.gov/ij/plugins/colocalization.html". Descriptive statistics and graphing was carried out using Microsoft Excel, and ANOVA and Pairwise multiple comparison T-tests (with Benjamini-Hochberg correction), Pearson correlation was conducting using R v3.4.1. Significant differences were considered when the P value < 0.05.

2.2.2.2.3 APBs assay

Following IF-FISH protocol in 2.2.2.2.1, slides were screened using the Olympus Cell^R/scan^R system to capture images of 100s-1000s of nuclei. Detection of APBs was performed following the criteria in Henson *et al.*, 2005. The analysis of co-localisation events was conducted on the captured images using ImageJ v1.5e with java plugin tool written by Pierre Bourdoncle.

2.2.2.2.4 Micronuclei analysis

Interphase nuclei were stained with DAPI, and 200-400 nuclei were manually counted. The micronucleation percentage was calculated as the following

$$\% \text{ Micronucleation} = \frac{\text{Number of nuclei with micronuclei}}{\text{Total number of screened nuclei}} * 100$$

2.2.2.3 DNA extraction

DNA was extracted from fresh frozen dry cell pellets. The cell pellet was re-suspended in 250 μl for ~1 x 10^6 cells of 1 X SSC (15 mM sodium citrate, 150 mM NaCl) buffer. Then cells were lysed by adding an equal volume of lysis buffer (10 mM EDTA, 100 mM NaCl, 100 mM Tris-HCl pH7.5, 1% sarkosyl). RNA was degraded by adding RNase A (at final concentration 10 mg/ml) to the mixture for 20 minutes. Proteins were digested efficiently by adding Proteinase K (final concentration of 20 mg/ml) and incubated in a water bath at 55°C overnight (~16 hours). The degraded proteins were extracted by adding an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1, Sigma) and gently mixing to emulsify the aqueous and organic phases in phase-lock-tubes (Qiagen) followed by centrifugation at 13000 rpm for 7 minutes at room temperature to separate the phases. The aqueous phase was removed into a new tube and the organic phase was safely discarded. The DNA was precipitated by adding 0.1 volume of 2 M NaOAc (pH 5.6) and 2.5 volume of 100% ethanol. The precipitated DNA was transferred carefully into a 1.5ml Eppendorf tube and washed with an appropriate volume of 80% ethanol and the DNA was dissolved in an appropriate volume of 5mM Tris-HCl pH 7.5. The concentration of DNA was estimated using a NanoDrop 1000 Spectrophotometer (Thermo Scientific, UK), and stored at -80°C.

2.2.2.4 RNA extraction

RNA was extracted from fresh frozen cell pellets using a Maxwell semi-automated method with Maxwell® RSC simplyRNA Cells Kit (Promega), following the manufacturer's instructions.

2.2.2.5 Polymerase chain reaction (PCR) technique.

Polymerase chain reaction (PCR) was carried out using 10-50 ng of genomic DNA or 1-10 ng of plasmid DNA, 0.3 pmoles of each primer, 0.1 U/μl of Taq polymerase (Kapa Biosystem) and 1X PCR buffer (termed 11.1X PCR buffer). The final concentration of the components of the PCR buffer in each reaction were: 45 mM Tris-HCl pH8.8 , 11 mM Ammonium Sulphate, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 4.4 mM EDTA pH8.0, 1 mM of each dATP, dCTP, dGTP, dTTP, and 113 pg/ml BSA as described in Jeffreys et al. (1988). The PCRs were set up in a class II laminar flow hood to avoid contamination. The thermal cycling step was carried out using a Veriti 96-Well Thermal Cycler (Applied Biosystems).

2.2.2.6 Single telomere length analysis (STELA)

STELA technique was adapted from Baird et al. (2003). In a class II safety laminar flow hood, the stock dilution of genomic DNA (10ng/μl) was diluted into 250pg/μl in H$_2$O containing tRNA (1ng/μl), and 0.02 pM Telorette2 linker before adding to a PCR master mix. PCRs were carried out in 10-25 μl volumes. Each 10μl reaction contained 0.9μl 11.1X PCR buffer, 0.3μM chromosome- or allele-specific primer, 0.01μM Teltail primer, 0.4U Taq (KAPA Biosystems), and 0. 25U PWO (Genaxxon Bioscience), 10ng tRNA, and 1μl DNA/tRNA/Telorette2 mix (final DNA concentration of 250pg-1000pg). The reactions were cycled as follow: 25 cycles of 96°C for 20 sec, 63 °C (XpYp E2, 17p6) 63 °C (12q-STELA) or 66.5°C (allele-specific primers XpYp-427G/415C) for 30 sec and 68 °C for 10 minutes. Each reaction was split into three identical 0.8% LE agarose gels for subsequence variant repeat detection following production of a Southern blot.

2.2.2.7 Agarose gel electrophoresis

Agarose gel electrophoresis was conducted using 0.5X TBE (89 mM Tris, 89 mM Boric Acid, 2 mM EDTA, and 0.5 pg/ml ethidium bromide). LE or HGT agarose (supplied by SeaKem or Lonza) were used. The agarose concentration varied from 0.8-2% depending on the size of DNA fragment to be resolved. LE agarose gels (1-1.5%) were used in most cases, but HGT with the concentrations from 1.5 to 2% were used to resolve digested PCR products (200- 500 bp) as stated in the T7 endonuclease I assay. DNA was visualised and gel images captured using a UV transilluminator (Syngene, Cambridge, UK).

2.2.2.8 Southern blotting

Agarose gels were processed in a depurinating acid solution (0.25M HCl) for 7 minutes, alkali-denaturing solution (0.5M NaOH, 1M NaCl) for 20 minutes, and in a neutralisation solution (0.5M Tris, 3M NaCl pH7.0) for 20 minutes with gentle shaking. The products were transferred onto 0.45μm nylon membranes pre-soaked in 3X SSC via capillary transfer that results from placing the gel onto 20X SSC soaked 3mm Whatman paper that was partially submerged in a reservoir of 20XSSC. Paper towels and an appropriate weight was added on the top of the blotting apparatus, and the paper towels were changed regularly on the first two hours to maintain consistent capillary action.

Mostly, the transfer was carried out overnight (~16 hours) to obtain fully transferred DNA onto nylon membrane. The membrane was dried in an 80°C oven for 10 minutes and cross-linked by exposing the blot to 700kJ per cm^2 of short wavelength ultraviolet light using the CL-1000 UV Cross-linker (UVP).

2.2.2.9 Probe labelling and hybridisation

The DNA transferred onto a nylon membrane by Southern blotting or dot blot method was detected using a radioactively labelled probe. Random-primed labelling of double-stranded DNA was used to generate α-^{32}P radioactively labelled DNA probes. This approach consisted of using random hexamers and the incorporation of [α-^{32}P]dCTP via the polymerase activity of the DNA Klenow fragment (Feinberg and Vogelstein, 1984). The labelling reaction was carried out in 30 µl volume. Approximately 10-15 ng of PCR amplified DNA or commercially bought DNA marker was denatured by boiling for 4 minutes before mixing with 6µl of 5X oligo-labelling solution (containing appropriate concentrations of dATP, dGTP, dTTP), and 1.5 [α-^{32}P]dCTP (0.555MBq), 12µg bovine serum albumin (BSA), 7.5U DNA polymerase I large Klenow fragment. The reaction was incubated overnight at 37 °C, and stopped by the addition of 1.7 volume of oligo-stop solution (20 mM NaCl, 20 mM Tris-HCl pH 7.5, 2 mM EDTA and 0.25% SDS). The probe was NaOAc-ethanol-precipitated via adding 900 µg of high molecular weight salmon sperm DNA as a carrier, and it was re-dissolved in 500 µl dH$_2$O. The labelled probe was denatured by boiling for 5 minutes before use.

Southern blots were pre-hybridised for ~20 minutes in phosphate buffer at 65°C prior to overnight hybridisation. In STELA variant repeat analysis, each membrane was incubated with a different variant repeat probe: Telomere /T-type probe (TTAGGG)$_n$, C-type probe (TCAGGG)$_n$, or G-type probe (TGAGGG)$_n$ (produced in the Royle's Laboratory), in combination with a mixture of probes to detect the 1kb + high range ladders. Pre-hybridisation and hybridisation were carried out in modified Church buffer (0.5M sodium phosphate buffer, 1mM EDTA pH 8.0, 14% SDS) at 65°C. The blots were then washed 3 X 20 minutes in high stringency buffer (0.1XSSC, 0.1% SDS). The blots were rinsed in water and kept damp in a saran film prior to overnight exposure to Amersham Phosphor Screens.

2.2.2.10 Telomere Restriction Fragment (TRF)

Approximately 6 µg of genomic DNA was digested with *RsaI* and *HinfI*, and resolved in a 0.6% agarose gel, but leaving telomeric DNA intact (Allshire *et al.*, 1989) and circular extrachromosomal telomeric DNA intact (Cesare and Griffith, 2004; Dlaska *et al.*, 2008). The digested genomic DNA was resolved in 0.8% of LE agarose. This followed by denatured hybridisation of telomere probe (TTAGGG)n to telomeric sequences.

2.2.2.11 C-circles assay

The C-circle protocol adapted from Henson et al. (2009) was used. For each cell line approximately 40 ng of digested genomic DNA was subjected to rolling circle amplification. The reaction included 1 mM of each (dATP, dGTP, dTTP, dCTP), with or without 7 U of phi29 DNA polymerase (from NEB), without a primer. The reaction was incubated at 30°C for 8 hours, followed by heat inactivation of the phi29 polymerase at 65 °C for 25 minutes. For quantitative analysis of the C-circle assay, each reaction was split into two equal volumes; one half was diluted with 2×SSC and dot-blotted onto a nitrocellulose membrane for native hybridisation with α-^{32}P -telomere probe (TTAGGG)$_n$, and the other half was denatured in a NaOH solution prior to 'denatured' dot-blot preparation and hybridisation to a α-^{32}P-Alu probe. C-circle abundance was obtained by normalising the signal of the telomeric probe (native hybridisation) to the signal from the Alu probe (denatured hybridisation) using ImageQuant software.

2.2.2.12 T7 endonuclease I assay

T7 endonuclease I can cleave DNA heteroduplex, as first shown by Mashal et al. (1995). This property of the T7 endonuclease I forms the basis for the mutation screening used to detect Indel mutations that can arise following CRISPR-Cas9 induced NHEJ (Figure 2.1). DNA was extracted from 1-2 x 10^5 cells. PCR primers were designed to amplify 500-600 bp amplicon that uniquely flanking the Cas9-gRNA cleavage site. PCR was carried out as described in describe above. A 3 µl volume of PCR products was size separated in a 2% LE agarose gel for PCR quantification with low molecular weight ladder. For the denaturation and re-annealing reaction, ~ 200 ng of PCR products were added to 1 x NEB buffer 2 in dH$_2$O. The reaction was denatured at 95°C for 5 minutes, then re-annealed in two ramping steps: 95°C - 85°C at -2°C/sec,

85°C-25C° at -0.1 C°/sec. The DNA heteroduplex was cleaved by adding 2.5 U of T7 endonuclease I (NEB) to the reaction and incubated at 37°C for 30 minutes. The whole reaction was then resolved in a 2% HGT agarose gel (SeaKem , Lonza) with a low molecular weight ladder for mutation detection analysis.
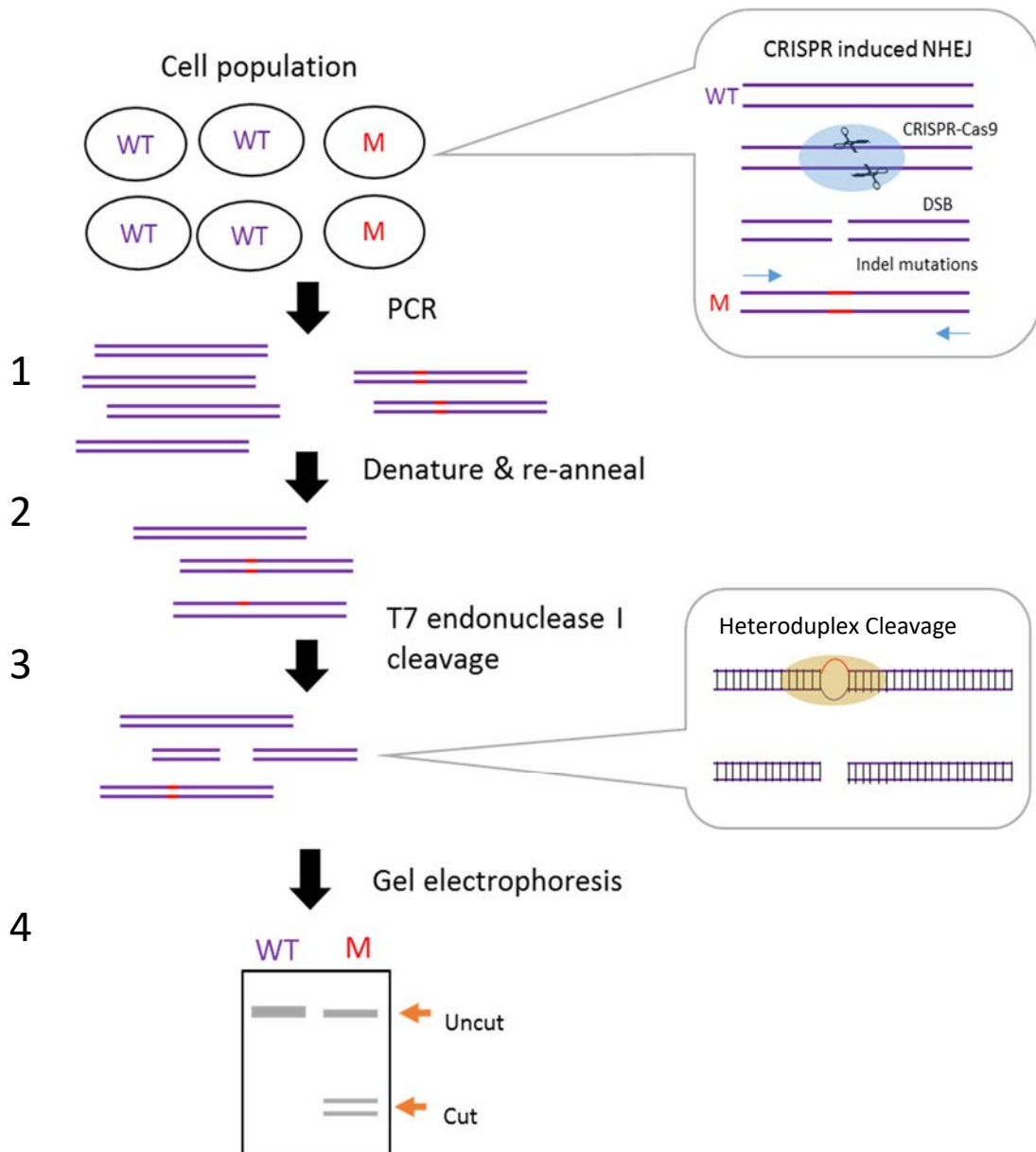
Figure 2.1 Schematic representation of the T7 endonuclease I assay. The assay is conducted in four steps: (1) PCR amplification of a region overlapping the CRISPR/Cas9 editing site, (2) Denaturing and re-annealing of amplified products, (3) Cleavage of heteroduplex molecules by T7 endonuclease I. (4) Gel electrophoresis. Uncut indicated WT and no CRISPR/Cas9 editing, and Cut indicated successful

detection of a mutation that arose from an error in DSB repair by NHEJ pathway. WT= wild type, M= mutated, NHEJ= non-homologous end joining.

2.2.2.13 Sanger sequencing – level 3.1

Sanger sequencing is a method based on the inclusion of chain-terminating dideoxynucleotides that lack a 3'-hydroxyl residue, during *in vitro* DNA synthesis as developed by Sanger et. al. in (1977). For the sequencing reaction, 1 µl of the Big Dye™ Terminator (version 3.1), and 3.5 µl of 5x Big Dye™ Terminator buffer were mixed with 20 ng per kb of PCR amplicon (or 70 ng per kb for Plasmid DNA), and 3.2 µM of the sequencing primer in a 20 µl reaction volume. The reaction was cycled 25-30 times as follows; 10 sec at 96°C, 5 sec at 50°C and 4 minutes at 60°C. The sequencing reaction was then cleaned to remove excess fluorescently labelled dideoxynucleotides by adding 2 µl of 2.2% SDS and incubation for 5 minutes at 98°C and 10 minutes at 25°C. Excess of Big Dye was removed by passing the reaction thought spin DTR gel filtration cartridge (Edge Biosystems). Gel running of the sequencing reaction was conducted by the Protein and Nucleic Acid Laboratory (PNACL) facility, University of Leicester. MacVector v15.5 and FinchTV v1.4.0 were used for sequence analysis.

2.2.3 Next generation sequencing

2.2.3.1 RNA-Seq

2.2.3.1.1 QC analysis and library preparation of RNA-seq

RNA purity was assessed using a Nanodrop (OD260/OD280) and an Agilent 2100 Bioanalyser with a Nano-RNA 6000 Chip was used to assess the integrity of the RNA. After the QC procedures, 2.5µg of each RNA sample was sent to Novogene Company Limited (Hong Kong) for library preparation and 12GB of Paired end (PE) non-stranded mRNA sequencing using an Illumina-HiSeq 4000. The mRNA library was prepared using an Illumina TruSeq protocol. This included mRNA poly-A selection with oligo (dT) beads, then the mRNA was fragmented randomly, followed by cDNA synthesis using random hexamers and a reverse transcriptase for the first-strand cDNA synthesis. The second strand was synthesised by nick-translation using an Illumina buffer that

contained dNTPs, RNAase H, and DNA polymerase I from *E. coli*. The double-stranded cDNA was phosphorylated and A-tailed, sequencing adapters ligated, and size selected prior to PCR amplification.

2.2.3.2 Expression profiling and differential expression analysis

Sequencing the cDNA libraries on an Illumina HiSeq 4000 resulted in >30 million PE reads for each sample. All data analyses were preformed using the available packages in high-performance computing (HPC) (ALICE2) at the University of Leicester.   The RNA-seq FASTQ files were assessed before and after removing the adaptors using

FASTQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After that clean reads were aligned to the human genome assembly (Homo_sapiens.GRCh38.82) using the TopHat2 algorithm and the default parameters (Trapnell *et al.*, 2012). Read counts per gene were calculated by Rsubread.v.1.26 (Liao *et al.*, 2013) using the Homo_sapiens.GRCh38.82 gene annotation. Normalisation and differential expression (DE) analysis were calculated by DESeq2 v1.16.1 (Love *et al.*, 2014). Fragments Per Kilobase of transcript per Million mapped reads (FPKM) method was used to obtain the normalised gene expression (Trapnell *et al.*, 2010).

2.2.3.3 Pathway enrichment analysis

Gene ontology (GO) terms and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathways associated with each DE gene were characterised by multiple enrichment tests (Benjamini-Hochbergi corrected, $p < 0.05$) using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang da *et al.*, 2009). The enrichment results were summarised by removing redundant GO terms (allowed similarity=0.7) using REViGO online tool (Supek *et al.*, 2011).

2.2.3.4 Whole Exome Sequencing (WES) analysis

2.2.3.4.1 QC analysis and library preparation of WES

Two methods were used for QC analysis of the DNA samples: (1) agarose gel electrophoresis was used to detect evidence of DNA degradation; (2) Nano Drop 1000 spectrophotometer (OD260/OD280) ratios were used as an indicator for DNA purity. After the QC procedures, 1μg of genomic DNA per sample was sent to Novogene

Limited (Hong Kong) for library preparation and 6GB of PE of WES using a Illumina-HiSeq 4000. At the Novogene sequencing facility, samples were re-quantified using a Qubit v.2.0 (for precise DNA concentration measurements). An Agilent SureSelect Human All ExonV6 kit (Agilent Technologies, CA, USA) was used for sequencing library preparation, and different index codes were added to each sample. Each genomic DNA sample was fragmented into 150-200bp fragment lengths by hydrodynamic shearing system (Covaris, Massachusetts, USA). DNA fragments with overhangs were converted to into blunt ended molecules with 5' -phosphorylated ends using a specialised repair reaction .This was followed by 3' end adenylation of the DNA fragments and oligonucleotide adapter ligation. Fragments with ligated SureSelect adapters on both ends were selectively enriched in a PCR reaction. Captured libraries were enriched in a PCR reaction to add index tags to prepare for hybridization. Finally, the prepared libraries were purified again using AMPure XP bead purification (Beckman Coulter, Beverly, USA) and quantified using an Agilent Bioanalyser 2100 system prior to sequencing.

2.2.3.5 Processing of the FASTQ files

The WES libraries were sequenced at 50X depth (>40 million PE obtained in each sample) on Illumina HiSeq 4000 platform. All data analyses were conducted using the available resources in HPC (ALICE2) at the University of Leicester. The adaptor sequences were trimmed from the FASTQ data using Skewer v0.2.2 (Jiang *et al.*, 2014). The FASTQ files were assessed before and after removing the adaptors using FASTQC v0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After that, free of adaptors FASRQ files were aligned to the human genome assembly (human_g1k_v37_decoy, from 1000 genome) using the Burrows-Wheeler Aligner (BWA v0.7.1) algorithm (Li and Durbin, 2009).The final Bam files were generated after several steps of converting Sam to Bam, sorting & indexing by SAMtools v1.3.2 (Li *et al.*, 2009), excluding the PCR duplicates by Picard v2.1.0 (Broad Institute, 2016), realignment and base recalibration by GATK v3.8 (McKenna *et al.*, 2010).

2.2.3.5.1 Single-nucleotide variant and Indel analysis

For each pair of normal-tumour samples, somatic Single Nucleotide Variant (SNV) calling was performed using three somatic callers SomaticSniper v1.0.5 (Larson *et al.*, 2012) (parameters: bam-somaticsniper -F vcf -G -L -q 1 -Q 40), Strelka v2.8.4 (Kim *et*

al., 2018) , and Mutect2 (Cibulskis *et al.*, 2013) (min-base-quality > 30). Somatic indels were detected by Mutect2. Intersection of SNVs was performed by BEDtools v2.25.0 using VCF files generated by each of the three somatic callers (Quinlan and Hall, 2010). Only SNVs that were detected by at least two somatic callers were considered (Figure 2.2). Functional annotation of SNVs and indels was carried out by ANNOVAR (Wang *et al.*, 2010). The statistical analysis of most frequently mutated genes was performed by MAFtools v1.7.05 (a R package) (Mayakonda, A. & Koeffler, H.P., 2016).
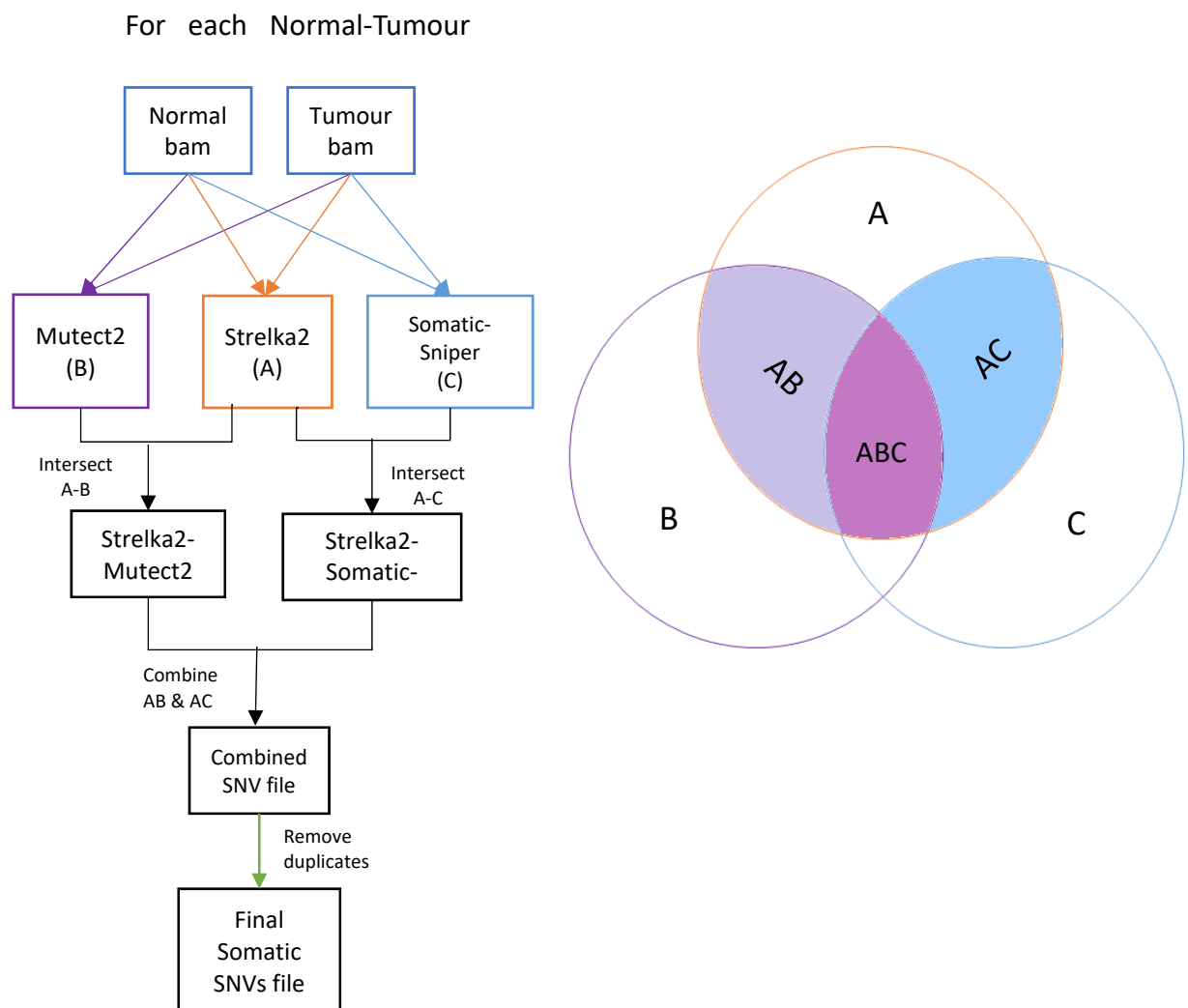
Figure 2.2 SNVs analysis pipeline using three somatic callers. Left, schematic representation of the pipeline that was used to assemble the list of shared SNVs that were detected by at least two SNV callers. Right, Venn diagram representation of the intersection of VCF files using BEDtools v2.25.0 (Quinlan and Hall, 2010).

2.2.3.5.2 Copy-number variant analysis

Allele-specific Copy Number Variant (CNV), Loss Of Heterozygosity (LOH) and estimation of ploidy analysis from the high-throughput WES sequencing was carried out using the CNV caller, FACETS v0.5.13 (Shen and Seshan, 2016), and also confirmed by Sequenza v 2.1.2 (Favero *et al.*, 2015). In FACETS, the analysis was carried out in two steps: (1) for each normal-tumour pair of Bam files, the single nucleotide polymorphism (SNP) arrays with read counts were collected from the exonic regions in the pileup format by SAMtools v1.3.2 (Li *et al.*, 2009) (parameters: mpileup -q15 -Q20 -B -f -r), (2) In R, the pileup file was automatically processed by joining the segmentation of total- and allele-specific read counts in order to calculate the minor-allele frequency and LOH. The copy number calls corrected for 'tumour purity', were then used for estimating changes in ploidy. The outputs were exported into genome view graphic plots and CNV data in tab-delimited files.

Chapter 3: The Roles of Orphan Nuclear Receptors in ALT

3.1 Introduction

Many studies have shown that ALT+ cells are characterised by a high rate of recombination at telomeres (Dunham *et al.*, 2000; Varley *et al.*, 2002). These recombination-like events of telomeres can change the telomere architecture and may contribute to spreading the variant repeats throughout the T-type (TTAGGG)n repeat array. The interspersion analysis of variant repeats with T-type using telomere variant repeat (TVR)–PCR at telomeres on chromosome 12q showed that C-type variant repeat was found to be present after ALT activation in all mutant 12qdeletion-haplotype telomere and not elsewhere in the progenitor map (Varley *et al.*, 2002). Thus, the following study using the whole genome sequencing as approached for variant repeat analysis showed that C-type variant was found to constitute ~ 14.6% of total telomeric repeats in WI38VA13_2RA cell line, while it constitutes ~ 2-4% of other screen ALT+ cell lines. Further, it was shown that WI38VA13_2RA has a lower abundance of G-type variant compared to C-type variant (Varley *et al.*, 2002; Conomos *et al.*, 2012; Lee *et al.*, 2014). Unexpectedly, the presence of a high density of C-type variant repeats of telomeres of the WI38VA13_2RA cell line was shown to bind to NR2C/F proteins (Dejardin and Kingston, 2009). They suggested that NR2C/F proteins may play a role in the ALT mechanism came from the detection of five of members of subfamily 2 nuclear receptors (NR2) with telomeres in ALT+, but not ALT- cell lines. These ORs are NR2C1 (also known as TR2), NR2C2 (TR4), NR2F1 (COUP-TF1), NR2F2 (COUP-TF2), and NR2F6 (EAR-2). Since then, other studies have shown that NR2F2 (COUP-TF2) and NR2C2 (TR4) proteins are associated with ALT markers and bound to existence of (TCAGGG)n repeats in telomeres of certain ALT+ cell lines (Conomos *et al.*, 2012; Conomos *et al.*, 2014; Marzec *et al.*, 2015). The consensus DNA binding site of the NR2F/C is RGGTCA(N0-8)RGGTCA, where 'N' is a nucleotide spacer ranging between 0 and 8 nucleotides, and 'R'=A/G (Figure 3.1).
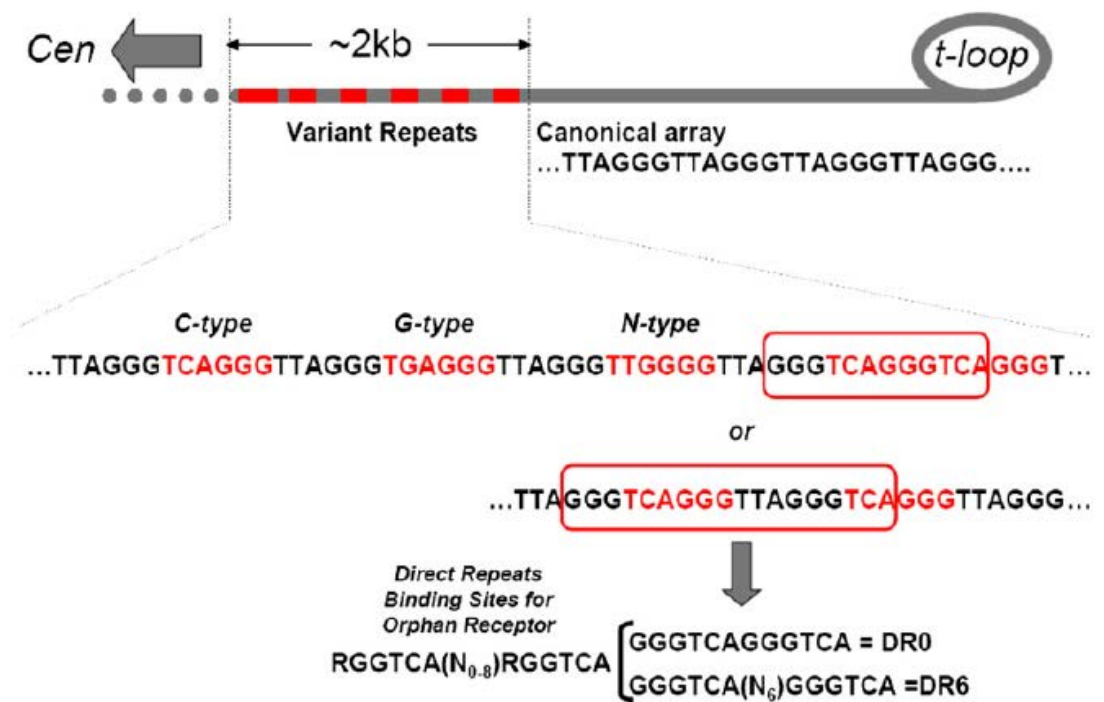
Figure 3.1 DNA recognition site of NR2C/F nuclear receptor proteins. This diagram shows the consensus DNA binding site of NR2C/F subfamily of proteins. In addition it shows potential binding sites for the NR2C/F proteins of telomeres that include arrays of (TCAGGG) variant repeats or these repeat interspersed with the canonical telomere repeat or other variant 6 mers (Dejardin & Kingston, 2009).

.

3.2 Aims

It has been shown that NR2C/F nuclear orphan receptors can bind to $(TCAGGG)_n$ repeats of telomeres of some ALT+ cell lines (Dejardin and Kingston, 2009; Conomos *et al.*, 2012; Conomos *et al.*, 2014), but it is unknown if all the telomeres in ALT+ cell lines can bind to these NRs in order to utilise the proposed mechanism to bring telomeres together to facilitate ALT. In this chapter, the expression of NR2C2 and NR2F2 and their the localisation at telomeres was investigated in a panel of five ALT+ cell lines (WI38VA13, WV, U2OS, SUSM-1, SAOS2) and one TEL+ cell line (HT1080). In addition, the abundance of potential NR2C2 and NR2F2 binding sites within selected telomeres was investigated in each cell line.

3.3 Results

3.3.1 The expression levels of NR2C2/NR2F2 proteins in ALT+ and TEL+ cell lines

NR2F2 and NR2C2 protein expression was measured by quantitative western blot analysis in relation to GAPDH, as a control, in the five ALT+ (WI38VA13, WV, U2OS, SUSM-1, SAOS2) and TEL+ (HT1080) cell lines. All three proteins were successfully detected at the expected molecular size. NR2C2 has two isoforms, 60-65 kDa; NR2F2 was detect as a single protein at 48 kDa; and GAPDH as a protein at 37 kDa (Figure 3.2 A). The proteins were quantified using ImageJ software, and two comparative analyses were made.

Data presented in Figure3.3B displays a comparison of NR2F2 and NR2C2 protein expression between cell lines. Significant variations in expression level of NR2F2, but not in NR2C2 were detected among the ALT+ cell lines. NR2F2 was expressed at a significantly higher level in the W-V cell line (~ 2 folds higher) compared to the other cell lines (WV vs HT1080, WI38VA13_2RA, U2OS, SUSM-1, SAOS2) (Figure 3.2 B). The next highest level of NR2F2 expression was observed in the WI38VA13_2RA cell line, significantly higher level than in HT1080, U2OS, SUSM-1, and SAOS2. It was also notable that NR2F2 was much less abundant in the SOAS2 cell line compared to all the others. No significant difference in the relative expression level of NR2C2 among ALT+ and TEL+ cell lines was detected (Figure 3.2 C).
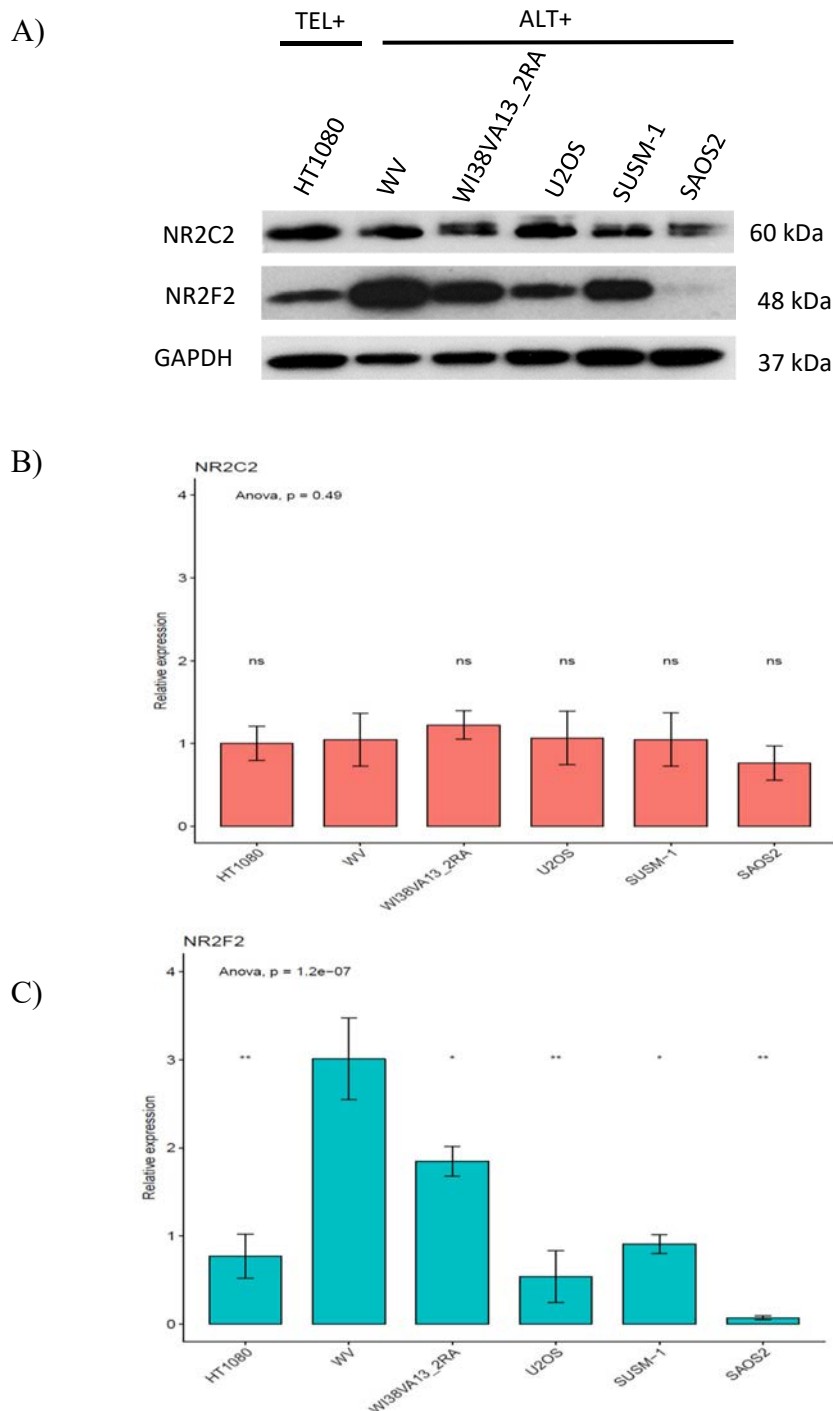
Figure 3.2 Quantification of NR2C2 and NR2F2 relative abundance in ALT+ and Tel+ cell lines. (A) Western blot detection of NR2F2, NR2C2, and GAPDH. (B) Quantification of the expression levels of NR2C2 relative to the loading control protein (GAPDH) (C) Quantification of the expression levels of NR2F2 relative to the control protein (GAPDH). WV was used as reference group in these comparisons. n= 3 independent experiments of biosocial replicates, *p < 0.05. **p<0.01. Error bars represent standard deviation (SD).

3.3.2 Comparison of NR2C2 and NR2F2 localisation at telomeres in the ALT+ cell lines

To measure the telomeric localisation of NR2C2 and NR2F2 in the selected panel of ALT+ cell lines, immunofluorescence (IF) detection of each them was combined with fluorescent *in situ* hybridisation (FISH) detection of telomeres on interphase nuclei (Figure 3.3 and Figure 3.4). NR2F2 and NR2C2 were detected by an Alexa488 dye-conjugated secondary antibody with maximum emission wavelength in 525 nm (green), and the telomeres by hybridisation to a peptide nucleic acid (Tel-PNA; Cy3 dye-$(CCCTAA)_3$) with maximum emission wavelength of Cy3 in 568 nm (red). The nuclei were counterstained with DAPI (4′,6-diamidino-2-phenylindole), maximum emission wavelength 470 nm (blue). To confirm the telomeric localisation of the NR2C/F proteins, the intensity profile analysis was carried out at some of the colocalised foci.WI38VA13_2RA was as positive control, and HT1080 as negative control (Figure 3.5). Subsequently, slides were screened using a scan^R Imaging system and large-scale analysis of co-localisation events was conducted on thousands of captured images using ImageJ colocalisation java plugin tool (written by Pierre Bourdoncle, Institut Jacques Monod, Service Imagerie, Paris). The frequency of telomeric localisation of NR2F2 and NR2C2 was calculated for each cell line. The differences in telomeric localisation frequencies for NR2C2 and NR2F2 were compared between the cell lines.
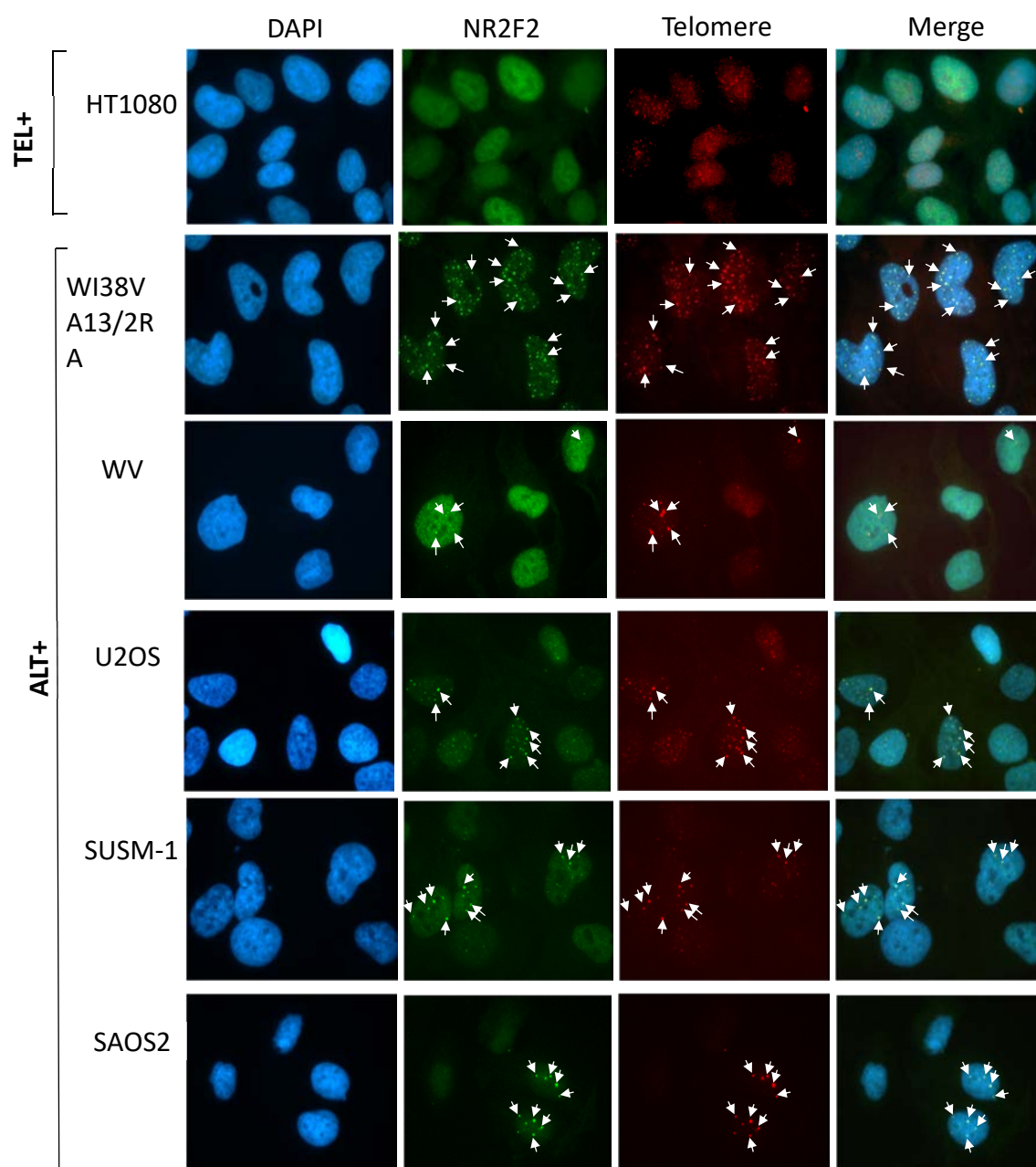
Figure 3.3 Detection of NR2F2 foci at telomeres in interphase nuclei of the ALT+ cell lines. IF staining for NR2F2 by Alexa 488 (green) and FISH detection of telomeres with Cy3–(CCCTAA)$_3$ (red) and nuclei counterstained with DAPI (blue). White arrows highlighted the colocalised foci.
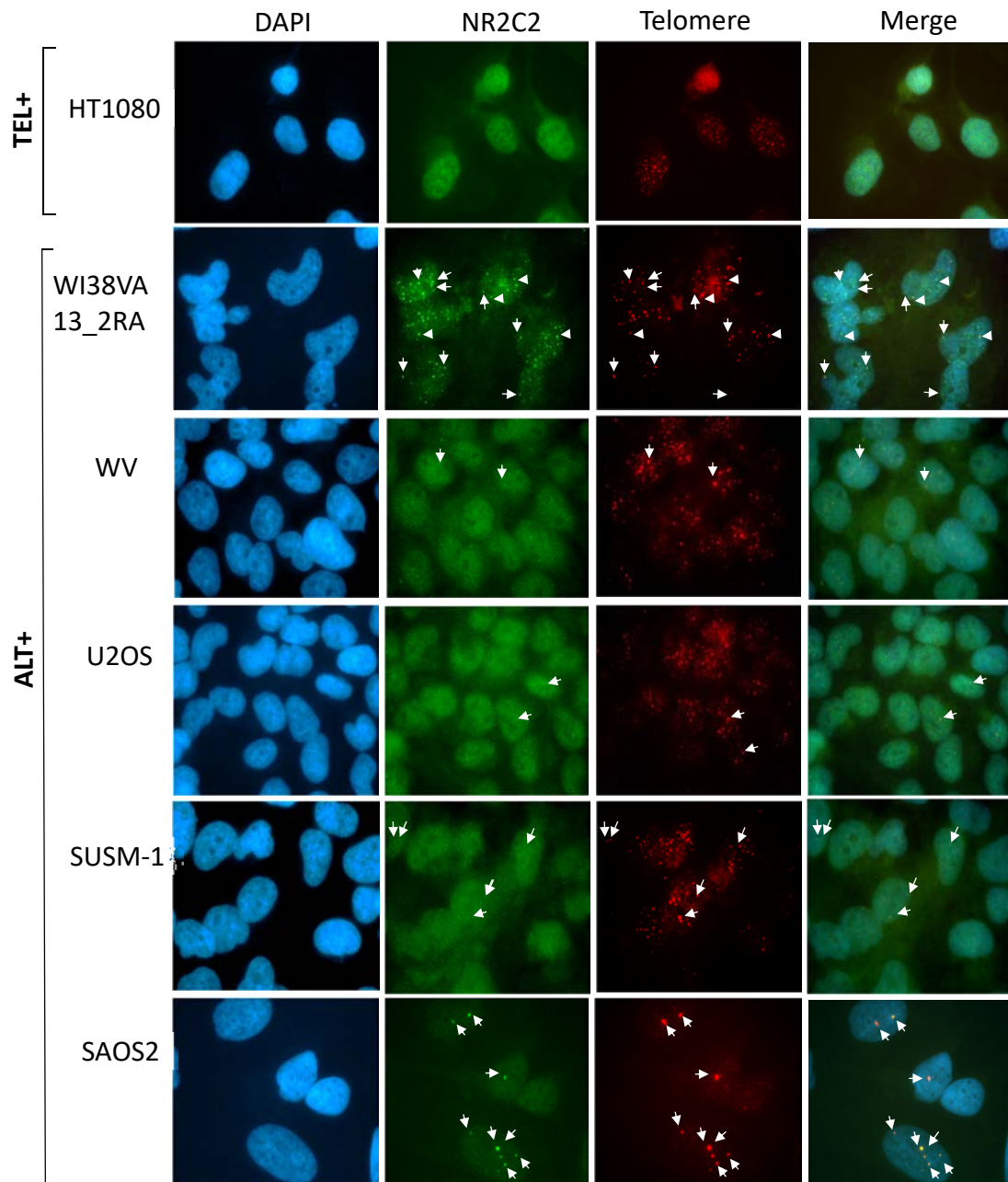
Figure 3.4 Detection of NR2C2 foci at telomeres in interphase nuclei of the ALT+ cell lines. IF staining for NR2C2 by Alexa 488 at interphase co-localised with Cy3–(CCCTAA)$_3$ telomeric probe (red) at interphase, and counterstained with DAPI (blue). White arrows highlighted the suspected colocalisation events.
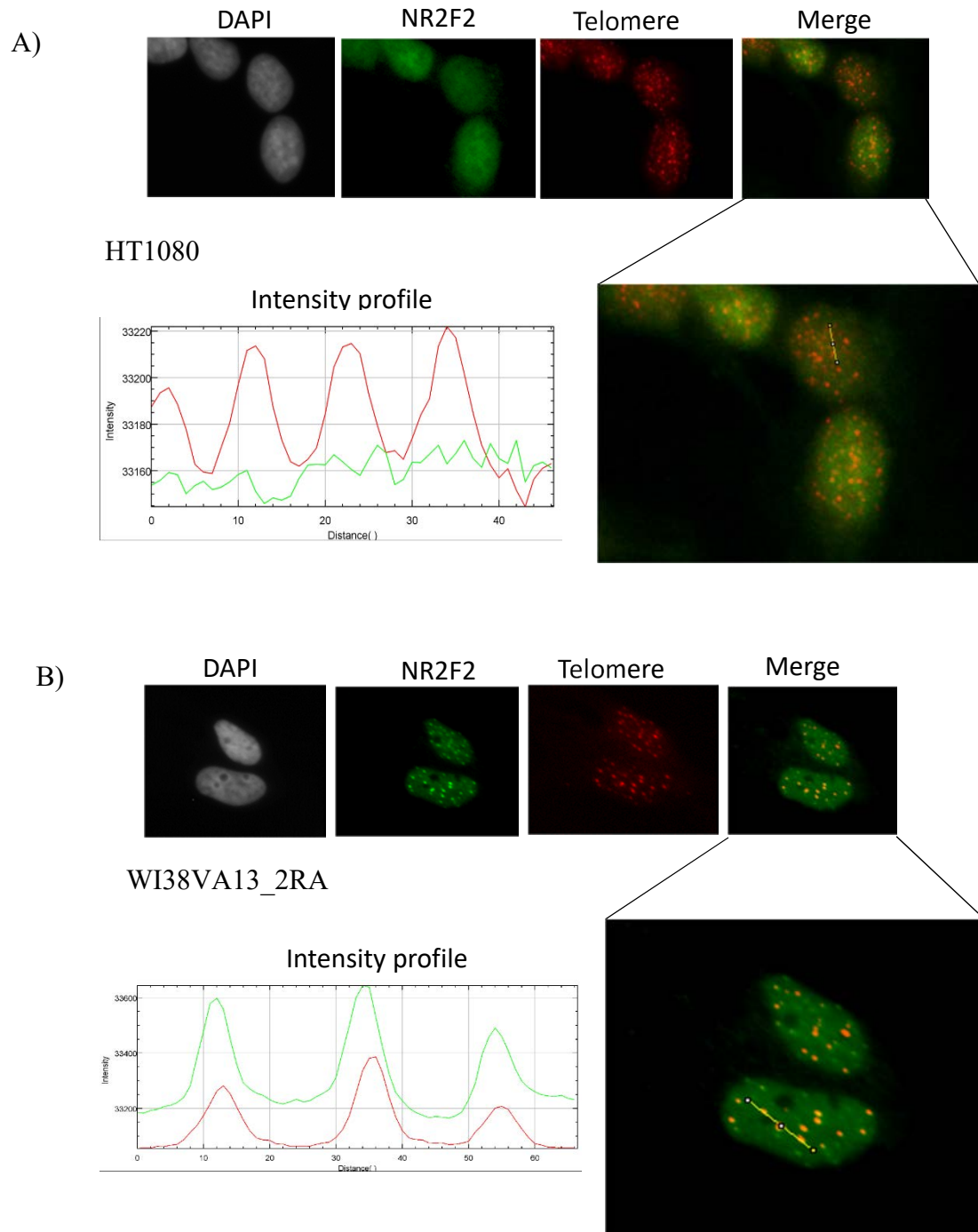
Figure 3.5 Inspection of colocalisation by measuring the intensity profile across foci. The plot multicolour function of ImageJ software was applied across selected foci in merged channels displays a plot of the intensity values along the line segment, NR2C2-Alexa 488 (green line), and Tel-Cy3 (red line). (A) HT1080 intensity profile plot showed high intensity with clear telomeric foci peaks but no NR2F2 peaks, which indicates no colocalisation. (B) WI38VA13_2RA intensity profile plot revealed overlapping peaks between telomeric foci and NR2F2 foci.

Data presented in Figure 3.6A shows NR2C2 - telomere localisation analysis. The percentage of NR2C2 foci localising to telomeres varied among ALT+ cell lines, ranging from the highest at 42% in WI38VA13/2RA to 0.7% in U2OS. The colocalisation frequency was significantly higher in WI38VA13_2RA compared to WV, U2OS, SUSM-1 ($p < 0.001$), and to SAOS2 ($p < 0.01$). Similar analysis with NR2F2 (Figure 3.6B) also showed that in the frequency of NR2F2 foci localising with telomeres varied among ALT+ cell lines, ranging from 91% in WI38VA13/2RA to 18% - 5% in other ALT+ cell lines. Similarly, colocalisation frequency of NR2F2 and telomeres was significantly higher in WI38VA13_2RA than WV ($p < 0.001$), U2OS and SAOS2 ($p < 0.0001$), and SUSM-1 ($p < 0.01$).

A)
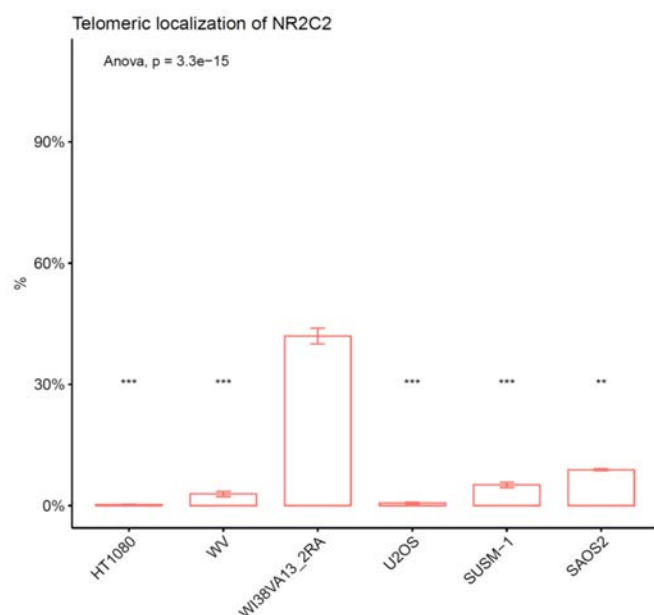
B)
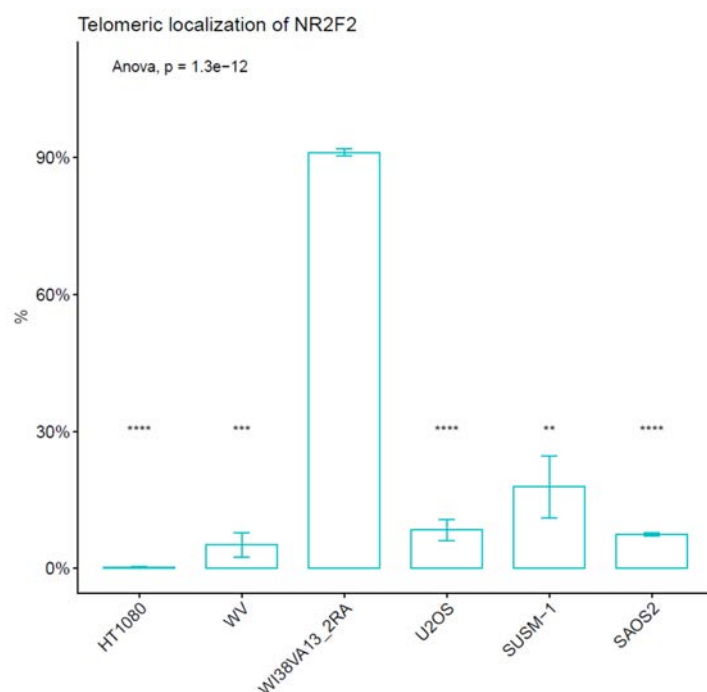
Figure 3.6 Quantification of the localisation of NR2C2 and NR2F2 foci at telomeres in ALT+ cell lines. (A) analysis of NR2C2 and (B) analysis of NR2F2 between cell lines. Frequencies of telomeric localisation were calculated by ImageJ software (mean ± SD, n= 3 independent experiments, number of counted cells > 5000 cells per cell line for each experiment). WI38VA13_2RA was assigned as reference group in (A-B). *p < 0.05, **p<0.01, ***p<0.001, ****p<0.0001.

3.3.3 Frequency of telomeres with the C-type variant in ALT + cell lines

3.3.3.1 Single telomere length analysis (STELA) for variant repeat detection

To investigate the relationship between the potential DNA binding sites for NR2F2 and NR2C2 with variation in the localisation of these proteins at telomeres, the abundance of the C-type variants was investigated in the same panel of selected cell lines. Single Telomere Length Analysis (STELA) was used to amplify telomere molecules from XpYp, 12q, and 17p using either chromosome or allele specific primers in the telomere-adjacent sequence from each cell line (Figure 3.7).

Each STELA product includes a constant length of telomere-adjacent sequence, which vary between chromosomes: these are 406bp, 544bp, and 3078bp when STELA amplicons are generates with the XpYp-E2, 12q-STELA, and 17p6 primers respectively. Different gel running time was used for a proper separation of telomere molecules generated by different chromosome specific primer. The size-separated STELA amplicons were analysed by Southern blot hybridisation to various α ^{32}P-dCTP labelled synthetic variant repeat probes (TTAGGG)n, (TCAGGG)n, and (TGAGGG)n for detecting T-type, C-type, G-type repeats respectively. Subsequently, measuring frequencies of telomeres with C-type (molecules that shown in C-type/G-type plots without any match in T-type plot was excluded from this analysis).

A)

B)

Figure 3.7 Schematic representation of the STELA technique used for detection of variant repeats. (A) STELA PCR. The telorette oligonucleotide has an 8nt sequence at the 3' end that anneals to the G-rich telomeric overhang. PCR is then carried out using chromosome or allele specific primer that anneals in the telomere-adjacent region, and a tetail primer that is the same sequence as the 5' end of the telorette. (B) An illustration of the internal structure of a STELA amplicon.

3.3.3.2 Frequency of XpYp telomeres with C-type variant in ALT + cell lines

The STELA XpYp-E2 chromosome specific primer (406 bp internal to the start of the telomere) was used to amplify the telomere molecules from both XpYp alleles. Replica blots were generated from each PCR and hybridised to probes at detect $(TTAGGG)_n$, $(TGAGGG)_n$ or $(TCAGGG)_n$ repeats. Analysis of the STELA blots showed that G-type repeats were detected in 87%, 37%, 57%, 67% and 53% of the XpYp telomere molecules in the HT1080, WI38VA13_2RA, WV, SUMS-1 and SAOS2 cell lines respectively but not in any of the U2OS XpYp molecules. C-type repeats were detected in a 76% in WI38VA13_2RA, and 47% in WV cell lines but not in any of the amplified molecules from the XpYp telomeres from the U2OS, SUSM-1 and SAOS2 cell lines (Table 3.1 and Figure 3.8). Detection of C-type repeats, which provide potential binding sites for NR2F2and NR2C2, in a high percentage of the XpYp STELA amplicons in the WI38VA13_2RA cell line is consistent with the published data (Dejardin and Kingston, 2009; Conomos *et al.*, 2012) and may reflect the inter-allelic processes that have increased the distribution of the C-type repeats in this cell line. However, C-type repeats were only detected in approximately half the XpYp STELA amplicons from the WV cell line and at a lower intensity. This suggested an allelic difference in this cell line, with one XpYp telomere allele containing the C-type repeats but the other not (Baird *et al.*, 1995).

Table 3.1 Frequency analysis of C-type and G-type contents at XpYp telomeres.

Cell line	Replicates	Chrom	T-type	C-type	G-type	% C/T	% G/T
HT1080	2	XpYp	124	0	108	0	87
WI38VA13_2RA	3	XpYp	254	192	93	76	37
WV	3	XpYp	237	111	134	47	57
U2OS	3	XpYp	192	0	0	0	0
SUSM-1	3	XpYp	189	0	126	0	67
SAOS2	3	XpYp	240	0	128	0	53
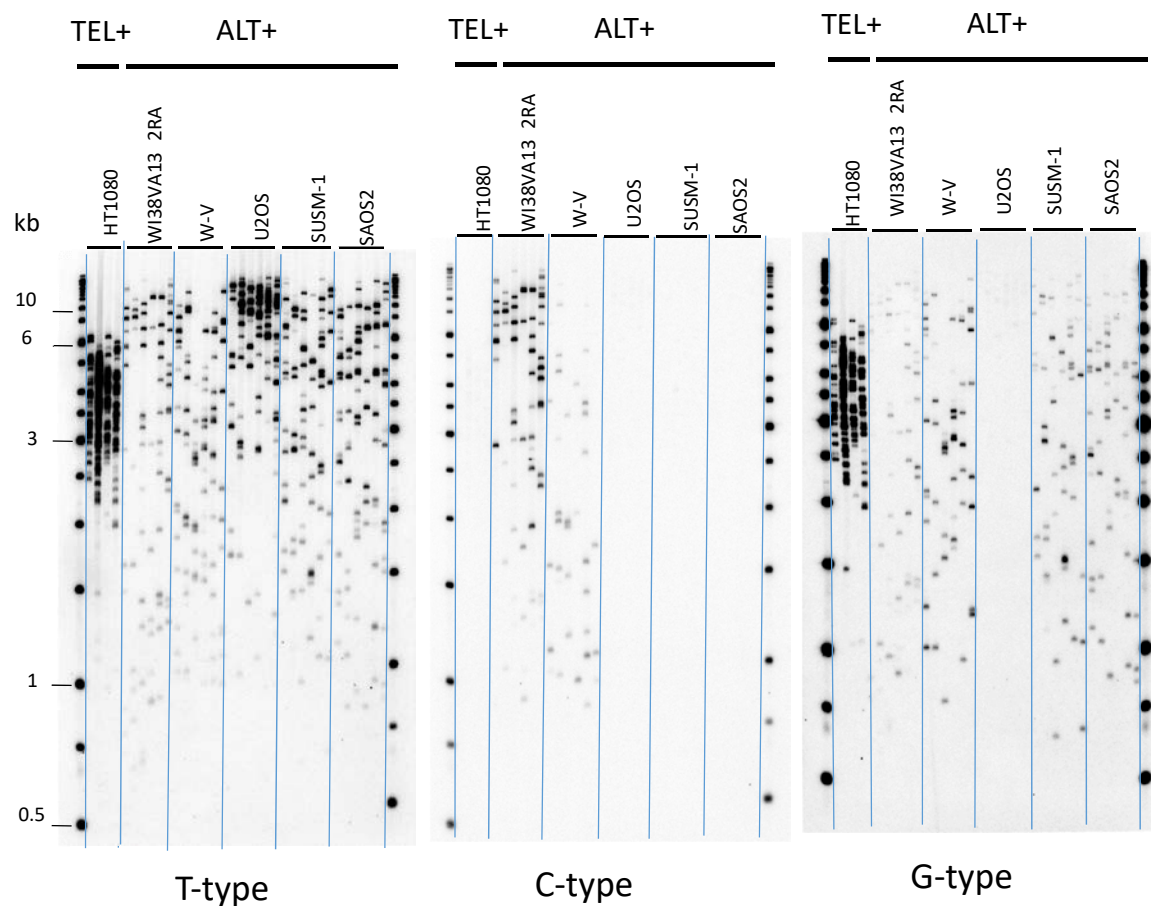
Figure 3.8 Analysis of variant repeats in XpYp telomeres. (A) Detection of variant repeats at XpYp telomeres. STELA was carried out using the XpYp chromosome specific primer (XpYp-E2) and the products were equally divided across three identical agarose gels. The three Southern blots were then hybridised to one probe to detect T-type, C-type, and G-type repeats.

To investigate if C-type content in WV cell line was linked to one XpYp allele, STELA was conducted using the allele-specific primers (XpYp-427G/415C and XpYp-427A/415T) that discriminate between alleles based on two SNPs in the telomere-adjacent sequence region (Baird *et al.*, 1995). Previous analysis had shown that the WI38 progenitor cell line is heterozygous for these XpYp SNPs but that the XpYp haplotype B associated telomere (amplified by the XpYp-427A/415T primer) has been lost from the WI38VA13_2RA cell line (Varley *et al.*, 2002). Therefore, WI38VA13_2RA was analysed side by side with WV as control for specificity of the allele-specific amplification. As expected, no STELA products were generated with the XpYp-427G/415C primer from the WI38VA13_2RA cell line, which validated the specificity of the PCR conditions (Figure 3.9).

For the WV cell, line the amplification efficiency was better with the XpYp-427G/415C than XpYp-427A/415T primer, and G-type repeats were only detected in the amplicons generated with the XpYp-427G/415C primer. The C-type repeats were detected by strong hybridisation signals in the amplicons generated by the XpYp-427A/415T primer, showing that this allele of the XpYp telomere in WV contains C-type repeats. There is a very weak signal with the C-type probe from the WV XpYp amplicons generate with the other allele specific primer XpYp-427G/415C. This could represent the presence of a very small number of C-type repeats in this allele or possible some non-specific binding of the probe.

In summary, this analysis supports the conclusion that there is an allelic difference between the distributions of C-type repeats in the XpYp telomere in the W-V cell line (Figure 3.9). This also was confirmed by sequencing molecules from XpYp and 16p fusions in WV cell lines. The analysis of telomere repeats at the XpYp side of fusion molecules showed that the C-type variant only in XpYp haplotype B, but not in XpYp haplotype A (Figure 3.10).
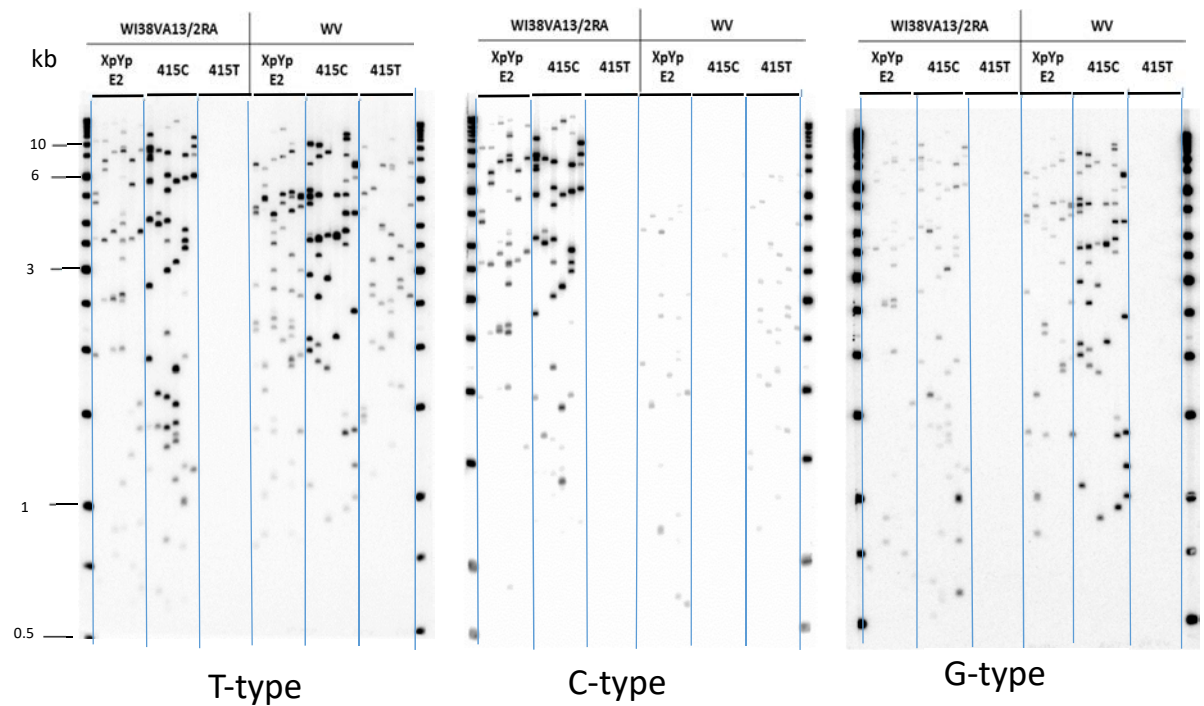
Figure 3.9 Assessment of variant repeat composition in alleles of the XpYp telomere in WV and WI38VA13_2RA. STELA was conducted using the chromosome specific (XpYp-E2) or allele specific primers XpYp-427G/415C, XpYp-427A/415T. The amplified products were equally divided between three identical agarose gels. Subsequently the three Southern blots were hybridised to different probes to detect T-type, C-type, and G-type repeats.
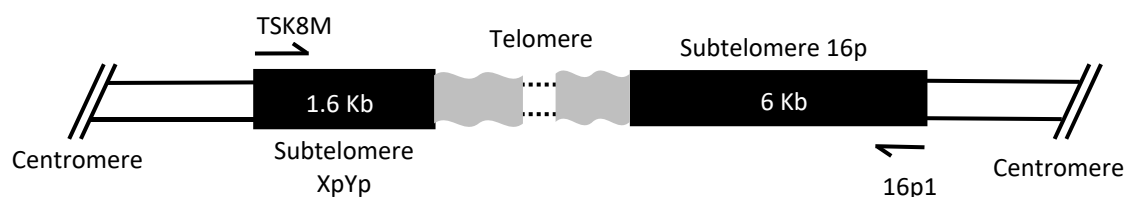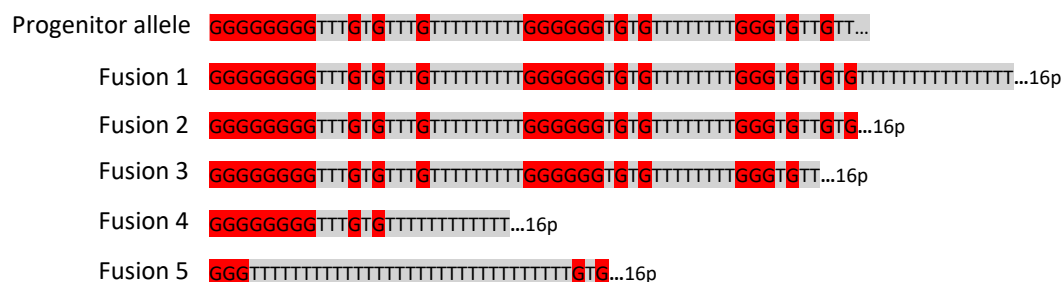
Figure 3.10 Variant repeats structure at XpYp in WV cell line. Above is schematic representation of XpYp and 16p fusion PCR. The verified fusion molecules were nested and sequenced. Single nucleotide polymorphisms (SNPs) at the telomere-adjacent region were used to subgroup aligned sequenced fusion molecules into their known XpYp haplotypes. The variant repeats structure of sequenced molecules from XpYp haplotype A were compared to their proposed progenitor allele that was annotated by TVR-PCR technique (Mendez-Bermudez *et al*, 2012).

3.3.3.3 Frequency of 12q telomere molecules with C-type repeats in ALT + cell lines

Previous studies have shown that the HT1080 cell line is homo (or hemi)-zygous for one haplotype across polymorphisms in the 12q telomere adjacent sequence (haplotype 12qB) and that the WI38VA13_2RA cell line is heterozygous for two haplotypes (12qΔ/12qB) (Varley *et al.*, 2002). The 12q-STELA primer (Figure 3.11) was used to amplify the telomeres from the same six cell lines to assess C-type and G-type repeat content.

Analysis of Southern blots generated in triplicate from 12q-STELA PCRs showed that C-type repeats could be detected in 12q telomeres of TEL+ and ALT+ cell lines, but at different frequencies; 79%, 80%, 33%, 40%, 39%, 34% in HT1080, WI38VA13, WV, U2OS, SUSM-1, and SAOS2 respectively (Figure 3.12 and Table 3.2). The amplicons showed different intensity profiles in between the cell lines following hybridisation to the C-type probe. The highest hybridisation intensity to the C-type probe was in WI38VA13_2RA, which reflected the high C-type content in these telomeres and the lowest C-type content detected in the SAOS2 12q telomeres. In this analysis, G-type was not detectable at telomeres of WI38VA13_2RA and SAOS2, while mean percentage of this variant was 72%, 8%, 51% and 40% in HT1080, WV, U2OS and SUSM-1 respectively. The low frequency molecules containing the G-type in the WV cell line may be a result of non-specific amplification from elsewhere in the genome, for example from 7q (Figure 3.12 and Table 3.2).

```
                      -550 SNP
12qa   TTTGGGGTAC CGTGGAGGCG AAGCAGCATT CTCCTCAGCA CAGACGCGGG
12qb   TTTGGGGTAC CGCGGAGGCG AAGCAGCATT CTCCTCAGCA CAGACGCGGG
                                          12qSTELA->
  7q   tttggggtac cacggaggcg aacgagcatt ctcctcagca cagacgcggg
12qΔ   .......... .......... .......... .......... ..........
```

Figure 3.11 Position of the 12q-STELA primer in the 12q telomere adjacent sequence. Sequence alignment of the telomere adjacent region of the three common 12q haplotypes (12qA, 12qB and 12qΔ) and the 7p telomere-adjacent sequence. The 12q-STELA primer anneals 544 bp from the start of the12q haplotype A and haplotype B associated telomeres but is absent from the flanking sequence of the 12qΔ-associated telomeres. Two bases are different between 7q and 12q telomere adjacent sequence at the 12q-STELA primer annealing site coloured in red. A SNP highlighted in yellow.
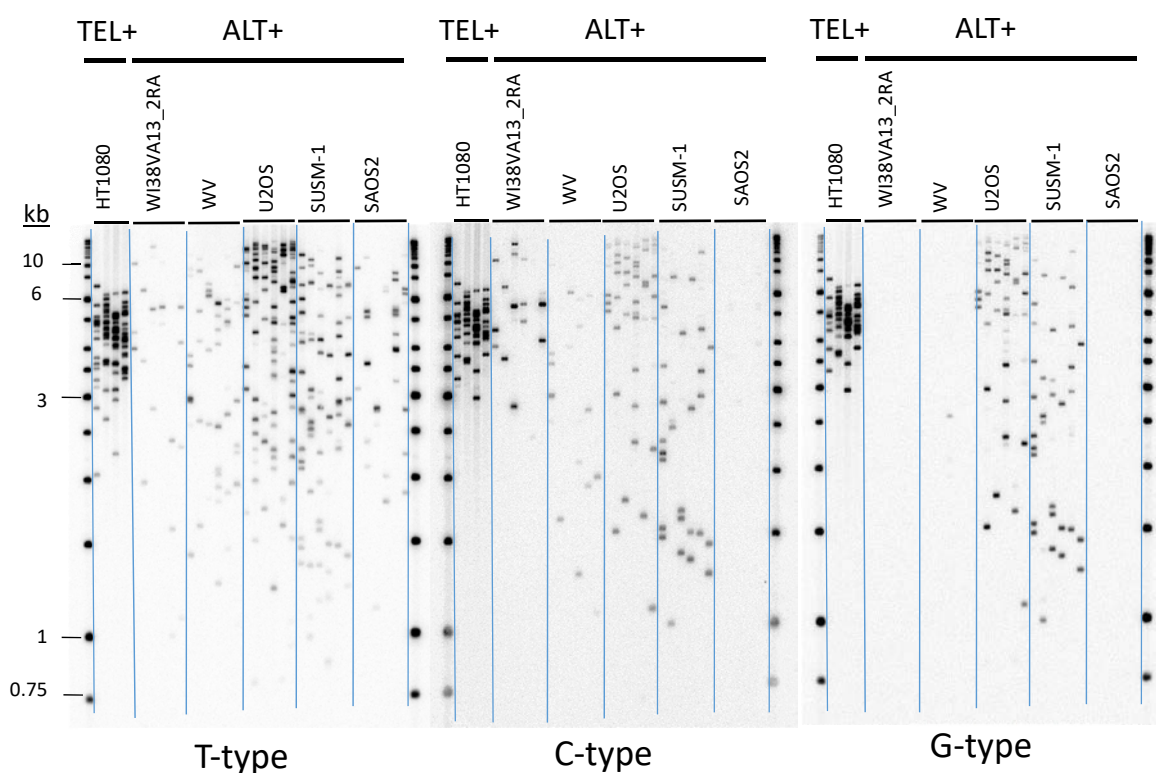
Figure 3.12 Analysis of variant repeats in 12q telomeres. Detection of variant repeats at 12q telomeres. STELA was carried out using 12q-STELA primer and the amplified products investigated in three identical Southern blots

Table 3.2 Frequency analysis of C-type and G-type contents at 12q telomeres.

Cell line	Replicates	Chrom	T-type	C-type	G-type	% C/T	% G/T
HT1080	2	12q	150	119	108	79	72
WI38VA13_2RA	2	12q	82	66	0	80	0
WV	2	12q	115	38	9	33	8
U2OS	2	12q	205	83	105	40	51
SUSM-1	2	12q	198	78	79	39	40
SAOS2	2	12q	85	29	0	34	0

3.3.3.4 Frequency of 17p telomeres with C-type repeats in ALT + cell lines

In this analysis, the 17p6 primer (3078 bp from the start of some 17p telomeres) was used to amplify the telomere sequences. PCR was carried out successfully in all cell lines, except SAOS2. It is known that some 17p telomeres (alleles) do not share the same subtelomeric sequence and so are not amplified by the 17p6 primers (NJR, personal communication) which may explain the failure to amply telomeres with this primer from the SAOS2 cell line.

Analysis of the Southern blots showed that the C-type repeats were not detected in the 17p telomeres in HT1080, WV, and SUSM-1 cell lines that suggesting a lack of binding sites for ORs. The C-type repeat was present in 90% of the WI38VA13_2RA and 74% of the SUSM-1 amplicons (Figure 3.13 and Table 3.3). The G-type repeat was detected only in a small percentage of 17p telomere molecules from SUSM-1. This might reflect movement of the G-type repeats by an inter-telomeric mechanism in the SUSM-1 cell line. Similar to what was found in XpYp and 12q telomeres, the 17p C-type repeat content was highest and most intense in the WI38VA13_2RA cell line, and with a lower frequency and intensity in 17p amplicons from U2OS.

Table 3.3 Frequency analysis of C-type and G-type contents at 17p telomere.

Cell line	Replicates	Chrom	T-type	C-type	G-type	% C/T	% G/T
HT1080	2	17p	77	0	0	0	0
WI38VA13_2RA	2	17p	125	113	0	90	0
WV	2	17p	259	0	0	0	0
U2OS	2	17p	95	70	0	74	0
SUSM-1	2	17p	169	0	22	0	13
SAOS2	2	17p	0	0	0	0	0
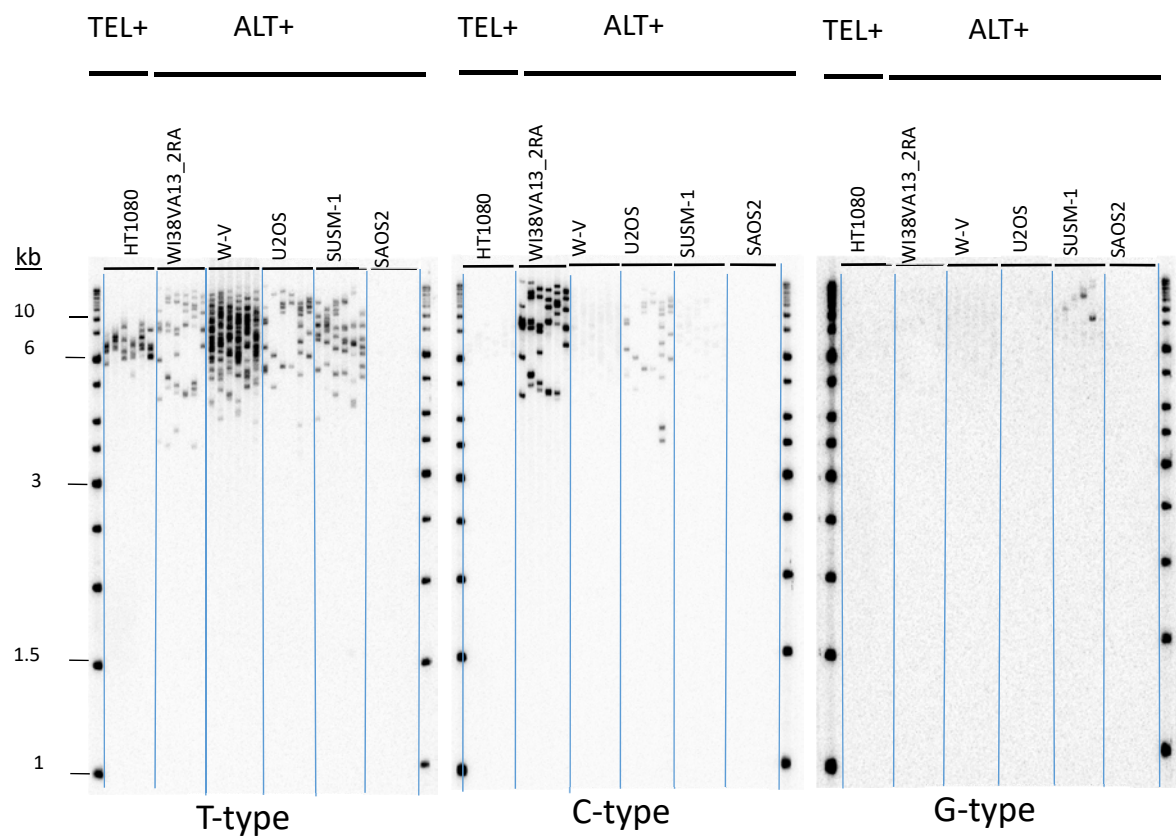
Figure 3.13 Analysis of variant repeats in 17q telomeres.. STELA was carried out using the 17p6 primer and downstream analysis conducted as for XpYp and 12q.

3.4 Discussion

3.4.1 Differences in protein levels and frequency of telomeric localisation of NR2F2 and NR2C2

Measuring the expression of the NR2F2 and NR2C2 proteins can be used to give an indication of whether they are required for the ALT mechanism. A previous study showed significant variation in the levels of NR2F2 and NR2C2 between ALT+ cell lines, but did not consider the variation in the level of these proteins between TEL+ and ALT+ cell lines (Conomos et al, 2012). In this work, the relative expression of the NR2C2 and NR2F2 proteins was measured in a panel of five ALT+ and one TEL+ cell line. Significant differences were found in the expression of the NR2F2 but not the NR2C2 protein among the ALT+ cell lines. Interestingly, NR2F2 was expressed at a significantly higher level in the W-V cell line (~ 2-fold increase) in comparison to other ALT+ cell lines, and was five times less abundant in the SAOS2 cell line.

The variation in protein levels between the ALT+ cell lines could arise in a number of ways: tissue origin and particular tissue specific roles; genome instability and loss of gene(s) following transformation; various roles in cancer pathways for example ALT. WV and WI38VA13_2RA are both SV40-trasnfromed fibroblast cell lines, but they demonstrated significantly different expression levels of NR2F2 (p<0.05), which could suggest they have a different requirement for NR2F2. Lower but similar levels of NR2F2 were found in U2OS and SAOS2, which are both osteosarcoma derived cell lines. However, NR2F2 was expressed at and much lower level in SAOS2 compared to U2OS that suggested the possibility that NR2F2 is more important for the biological pathways in U2OS than SAOS2. This suggest expression varies between osteosarcoma derived cell lines.

The genomes of ALT + cell lines are highly rearranged showing ploidy changes, reciprocal and non-reciprocal translocations and other rearrangements that can result in DNA copy number changes. Therefore, the relative abundance of NR2F2 arise from either gains or losses of functional copies of the gene (located at 12q26.2). If this underlies the expression changes then it, the NR2C2 gene (located at 3p25.1) might has a stable copy number and therefore expression level of NR2C2.

The telomeric localisation analyses showed that both NR2F2 and NR2C2 frequently colocalise with telomeres in the WI38VA13_2RA cell line, 91% and 42% respectively, in comparison to other ALT+ cell lines, consistent the with work previously conducted by Dejardin and Kingston (2009) and Conomos et al (2012). In contrast, the high level of NR2F2 expression in the W-V cell line did not result in preferential localisation at telomeres as only 5.13% of telomere foci were also positive for NR2F2. However, the high nuclear content of NR2F2 in WV cell line could mask some telomeric localisations that might not have been detected in this analysis (only NR2F2 foci that were more intense than the background level were included in downstream colocalisation analysis). The other ALT+ cell lines (U2OS, SUSM-1, and SAOS2) showed significantly lower NR2F2-telomere colocalisation frequencies and much lower for NR2C2. The correlation test for telomeric localisation and expression level of ORs in ALT+ cell lines (using Pearson method) showed no significant relationship, NR2C2 (p=0.31), NR2F2(p=0.43) (Figure 3.14). In summary, this does not support a significant role for these ORs in the ALT mechanism for most ALT+ cell lines. This is discussed further in Chapter 4.

A)
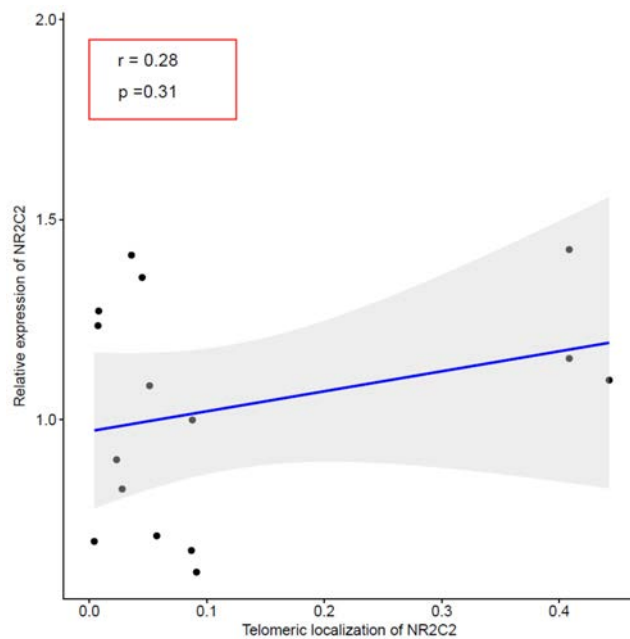
B)
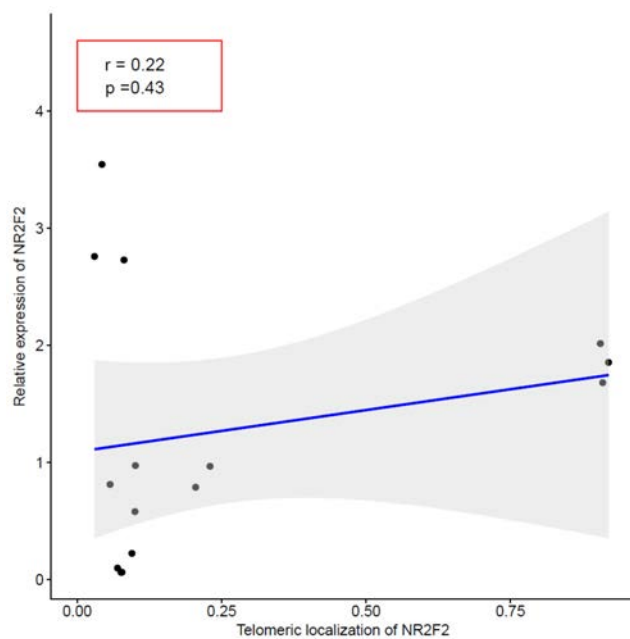
Figure 3.14 No correlation between the relative expression of NR2F/C proteins and their telomeric localisation was found for ALT+ cells. The correlation coefficient (r) and p value was calculated using the Pearson method. Shaded region in grey is a 95% confidence interval on the fitted values, where the true regression line lies within. (A) Correlation test between the relative expression and the telomeric localisation of NR2C2. (B) The relative expression and the telomeric localisation of NR2F2.

3.4.2 Absence of C-type repeats in some telomeres of ALT+ cell lines

Frequencies of telomere molecules with C-type repeats were measured at three telomere telomeres XpYp, 12q, and 17p using STELA and Southern blot hybridisation to $(TTAGGG)_n$ or variant repeat probes. In WI38VA13_2RA, high frequencies of telomeres with C-type were found: 76%, 80%, and 90% in XpYp, 12q, and 17p respectively. This is consistent with previous publications that showed no significant content of C-Type repeat at telomeres of the ALT+ cell lines except what has been reported in WI38VA13_2RA (Varley *et al.*, 2002; Conomos *et al.*, 2012; Lee *et al.*, 2014). However, C-type repeats were not detected in XpYp telomeres in the U2OS, SUSM-1 and SAOS2 cell lines. Similarly, the C-type variant was detected in 17p telomeres in WV and SUMS-1. These telomeres are unlikely to be able to bind to NR2F2, NR2C2 or other ORs with similar preferred binding sites, which is supported by the co-localisation frequencies reported in this chapter. Consequently, telomeres that lack $(TCAGGG)_n$ repeats are unlikely to act as donors or recipients of telomere lengthening via the proposed OR-ZNF817-NuRD mechanism.

However, other telomeres with potential NR2C/F binding sites can be present within the same cell. This could suggest that more than one mechanism for bringing telomeres together prior to ALT can be used by the same cell. For example, recent evidence has shown that telomeres can be brought together by various RAD51-dependent and independent mechanisms in the U2OS ALT+ cell line. Rad51 with heterodimers of the meiotic recombination factors, Hop2-Mnd1, appeared to be localised to APBs and telomeres clustering in ALT+ cells and to facilitate long-range telomere movement searches for homologous sequences (Cho *et al.*, 2014; Dilley *et al.*, 2016). This could be the mechanism for bringing telomeres together in ALT+ cells that lack the C-type repeats in some of their telomeres.

Chapter 4: Transient downregulation of NR2F2 by siRNA

4.1 Introduction

4.1.1 Role of NR2F2 in cancer pathways

NR2F2 acts as a transcriptional activator or repressor of a large network of 413 genes in mammalian cells (gene list from the Genecards database, January, 2018; Appendix 4.1). The targets include genes that code for twenty hormonal and inflammatory receptors, eighteen zinc-finger proteins, seventeen protein kinases, a group of growth and differentiation related proteins, seven histone clusters, seven sialic acid binding proteins, three protein phosphatases, five RAS family proteins, four transcription factors, four ATPases, four Rho GTPases, four small nucleolar RNA, two SWI/SNF-related, and others.

The role of NR2F2 in oncogenic and tumour suppressive functions has mainly been investigated in breast and prostate cancers. For example, in breast cancer it has been shown that NR2F2 works as a tumour suppressor by interacting with SMRT to repress PPARG expression and thus affect adipogenesis (Eubank et al, 2001; Okamura et al, 2009; Zheng et al, 2016). It has also been shown that NR2F2 can act as an oncogenic factor in four distinctive pathways in cancer. First, by activating the angiogenesis pathway, which is normally inactive after cell differentiation and following the embryogenesis. Expression of NR2F2 in prostate and pancreatic tumours seems to regulate the vascular endothelial growth factor (VEGF) gene family and their receptors. For example, depletion of NR2F2 in a pancreatic tumour model was associated with elevated VEGR-1, and downregulation of VEGFR-2 expression resulting in a significant reduction in angiogenesis (Qin *et al.*, 2010). Second, NR2F2 can play roles in cellular matrix and metastasis pathways. Expression of NR2F2 in breast cancer (MDA-MB-231) and lung cancer cell lines (A549, H520, and H441) resulted in elevating the expression of extracellular matrix-degrading proteinases, including the matrix metalloproteinase 2 (MMP2) and urokinase-type plasminogen activator (uPA) that are known to be associated with fibrinolysis, cell migration, and angiogenesis (Navab *et al.*, 2004; Annecke *et al.*, 2008). Third, by modulating the expression of cell cycle and cellular growth factors. For instance, NR2F2 controls

expression of the E2F transcription factor 1 (E2F1) in endothelial cells. E2F1 is an important player in regulating the G1/S transition (Johnson *et al.*, 1993). NR2F2 also plays a role in the Notch signalling pathway by direct regulation of downstream and upstream genes including Hairy and Enhancer of Split-Related Protein 2 (HEY2), Hes Family BHLH Transcription Factor 1 (HES1) and 1 (HES2), Nicastrin (NCSTN), Forkhead Box C1 (FOXC1), Neuronal Pentraxin 1 (NPTX1) and others (Chen *et al.*, 2012). Fourth, NR2F2 can contribute to evasion of cell senescence and the influence of growth suppressors. Elevated NR2F2 expression in prostate cancer is associated with SMAD4 in inactivating TGFB1/TGF-β and consequently in regulating TGFB1's downstream target genes (such as CDKN1A/p21 and CCND1/CyclinD1), thereby enabling the cell to bypass the TGF-β growth suppression barriers (Qin *et al.*, 2013).

4.1.2 NR2F2 targets and ALT mechanism

The intersection between the 413 NR2F2 regulated genes and the 73 ALT related genes (reviewed in section 1.4.3.5 in chapter 1) showed that only three genes overlap. These genes are TP53, RTEL1, and HOP2/PSMC3IP (Figure 4.1). TP53 plays major roles in a cell proliferation barrier and it is usually inactivated in cancer cells, facilitating the bypassing of the M1 checkpoint, contributing to genomic stability and apoptosis (Stagno D'Alcontres *et al.*, 2007). In the cell lines used in this work, the expression of the SV40 large T-antigen will inactivate the TP53 pathways in WI3VA13_2RA and WV. The U2OS cell line expresses wild type p53 protein but the p53-mediated induction of apoptosis is evaded by another route (Jiang *et al.*, 2015). Regulator of telomeres RTEL1 encodes an ATP-dependent DNA helicase that contributes in genomic stability and telomere length regulation. Ablation of RTEL1 in mouse embryonic stem cells resulted in inhibition of telomere extension, and sensitivity to a range of DNA damaging agents (Uringa *et al.*, 2012). RTEL1 also plays a role in telomere stability by dismantling T-loop and resolving telomeric G-quadruplex (G4) DNA structures in mouse embryonic fibroblasts (Vannier *et al.*, 2012). Thus, a change in RTEL1 expression is a potential route through which NR2F2 could influence telomeres in ALT+ cells.

HOP2/PSMC3IP encodes a meiotic recombination factor that is part of the HOP2-MND1 complex. It interacts with other proteins to stimulate RAD51-mediated strand exchange during meiosis. The HOP2-MND1 complex with RAD51 localised to APBs and facilitates the large-scale homology search in the recombination-based ALT mechanism (Cho *et al.*, 2014; Dilley *et al.*, 2016). Thus, a change in expression of HOP2 is predicted to have a major effect on telomere lengthen in ALT+ cells.
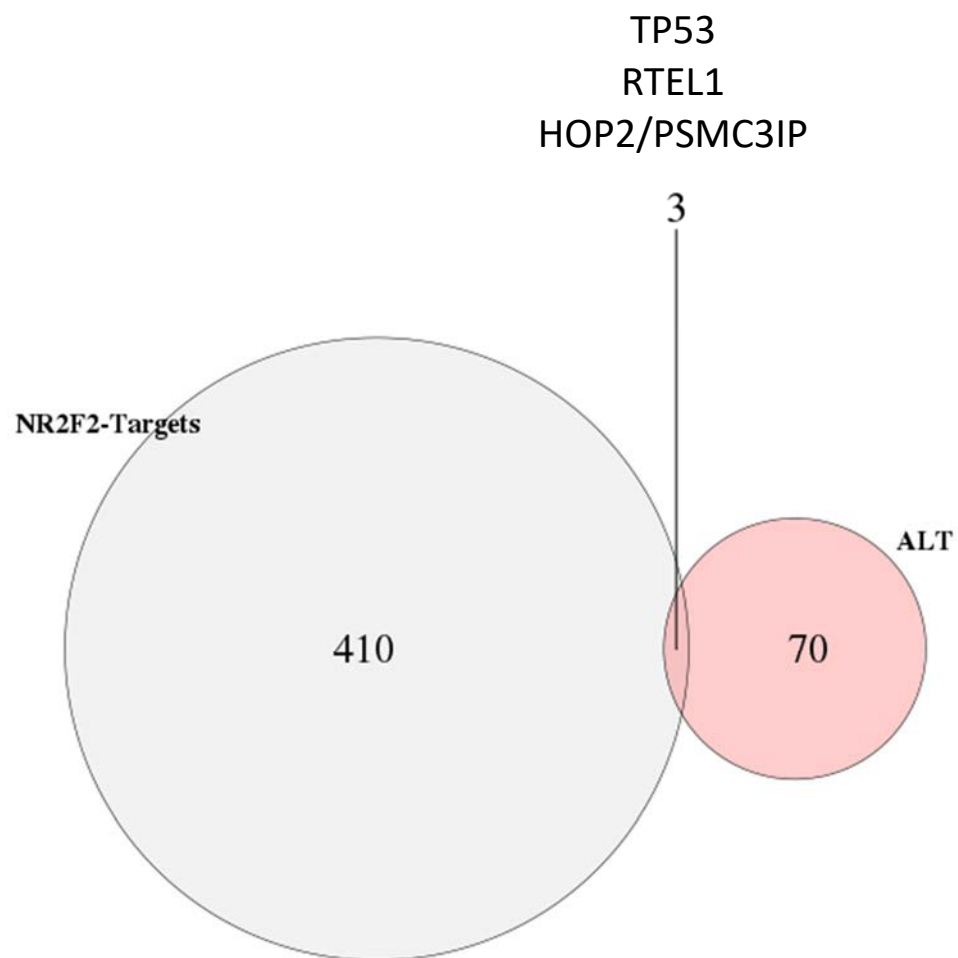
Figure 4.1 Venn diagram of the intersection between ALT-related genes and NR2F2-targets.

4.2 Aims

Depletion of NR2F2 in the ALT+ WI38V13_2RA cell line has been reported to result in a significant reduction in ALT markers (Dejardin and Kingston, 2009; Conomos *et al.*, 2012). In the previous chapter I showed low frequencies of telomeric colocalisation and the absence of C-type variant repeats (the potential binding site for NR2F2) in some telomeres of some other ALT+ cell lines suggesting that NR2F2 may not be involved in telomere lengthening in these cell lines. Furthermore, the elevated level of NR2F2 protein expression (~ 2-fold) in the WV compared to other ALT+ cell lines was not associated with significant localisation at telomeres. It is possible however, that NR2F2 is required for other biological pathways in these ALT+ cell lines. In order to investigate this, NR2F2 expression was downregulated using siRNA in WV and other selected ALT+ cell lines. This was followed by measuring changes in ALT markers, cell cycle and gene expression analysis.

4.3 Results

4.3.1 Selecting cell lines for NR2F2 siRNA

From the panel of cell lines that were used to investigate the expression and role of NR2C/F proteins in ALT mechanism in chapter 3, three were selected for transient NR2F2 downregulation. WI38VA13_2RA (SV40-transformed, female fibroblast cell line) was used as positive control as it has been shown that depletion of NR2F2 results in significant reduction in the number of APBs, but no change in c-circle abundance (Dejardin and Kingston, 2009; Conomos *et al.*, 2012). WV (SV0-transformed, derived from normal fibroblast of a 45-year-old male patient with Werner's syndrome) was selected because it showed significantly higher NR2F2 protein expression compared to the other ALT+ cell lines. U2OS (Osteosarcoma, female, with wild type p53) was chosen because it showed relatively low NR2F2 expression in compare to WV and WI38VA13_2RA. In addition, HT1080 (Fibrosarcoma, male, TEL+) was used as negative control for ALT markers in this work.

4.3.2 Assay for the effectiveness of siRNA at depletion level of NR2F2 protein.

NR2F2 protein domains have >90% similarity with NR2F1 protein domains (Litchfield and Klinge, 2012). To identify potential NR2F2 siRNA sequences that avoid targeting other genes, short sequence blast was performed against the human transcriptome. The selected anti-sense siRNA sequence showed 100% matched to NR2F2 exon 2 and no significant similarities (<60%) with other genes and it was predicted to target all five NR2F2 splice variants. Furthermore, multiple sequence alignment of NR2F2 and NR2F1 transcripts showed that the selected siRNA included 5 mismatches with NR2F1transcripts, indicating that it would be specific for NR2F2 (Figure 4.2).

A)

```
5'- CCACUCGUACCUGUCCGGAUAUAUU - 3'  Anti-sense
3'- GGUGAGCAUGGACAGGCCUAUAUAA - 5'  sense
```

B)
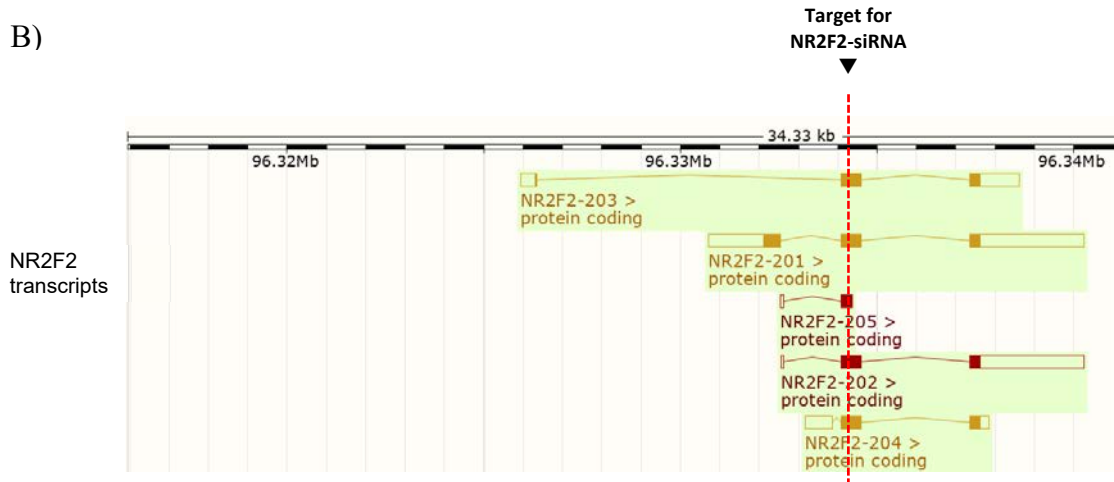
C)
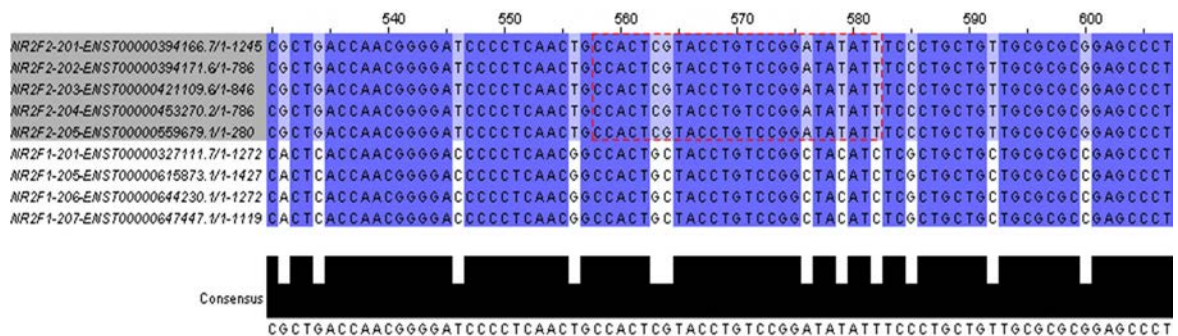
Figure 4.2 Design of the NR2F2 stealth siRNA oligonucleotide sequence on NR2F2 cDNA. (A) Sequence of the NR2F2 siRNA, (B) location of NR2F2 stealth siRNA target sequence in the five known splice variants, (C) position of the anti-sense NR2F2 siRNA target on alignment of NR2F2 and NR2F1 cDNA sequences showing five nucleotide differences. CLUSTAL Omega v1.2.4 was used to generate the multiple sequence alignment of cDNA sequences.

The stealth siRNA sequences were synthesised and annealed (Life Technologies) and delivered to the WV cells using Lipofectamine® RNAiMAX Reagent with Opti-MEM® Medium. To achieve efficient downregulation for NR2F2, two different quantities of stealth siRNA duplex were tested (25 pmols, 50 pmols). Cell pellets were collected 48 and 72 hours after transfection and western blot analysis was conducted to measure the level of NR2F2 down regulation. The maximum reduction in NR2F2 expression was observed after 72 hours using 50 pmols of siRNA (Figure 4.3). The results from this experiment were used as a guide to estimate the quantity of siRNA required to achieve depletion of NR2F2 protein in WI38VA13_2RA and U2OS cells (based on the NR2F2 expression relative to the WV cell line). Measuring of NR2F2 expression in these three cell lines after 72 hours of NR2F2 siRNA treatment showed reduction in NR2F2 level by ~ 87% in WV, ~84% in WI38VA13_2RA , ~ 94% in U2OS compared to the control siRNA (Figure 4.4).
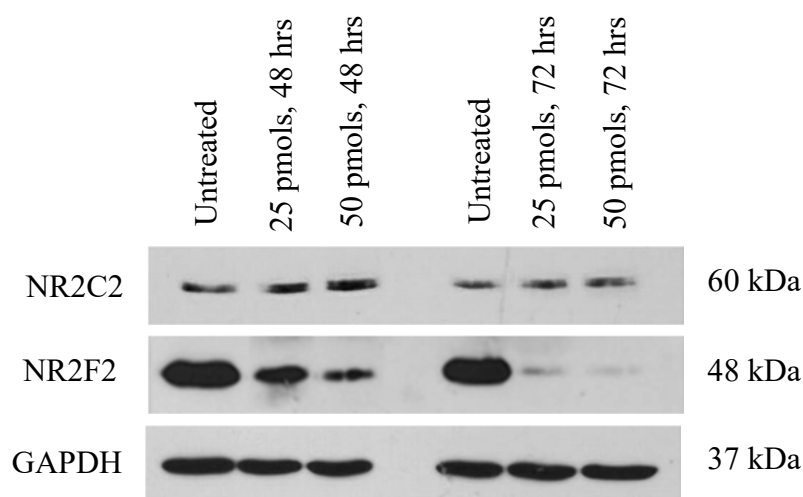
Figure 4.3 Optimising the siRNA mediated depletion of NR2F2 using the stealth siRNA. Western blot detection of NR2F2, NR2C2, and GAPDH proteins. The amount of siRNA used together with the time after transfection are given. This experiment has carried out using WV cells.

A)
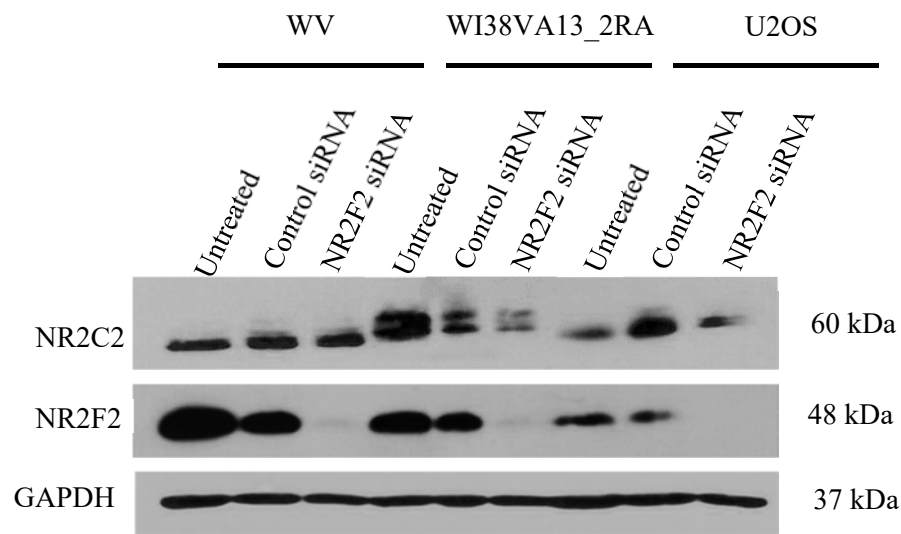
B)
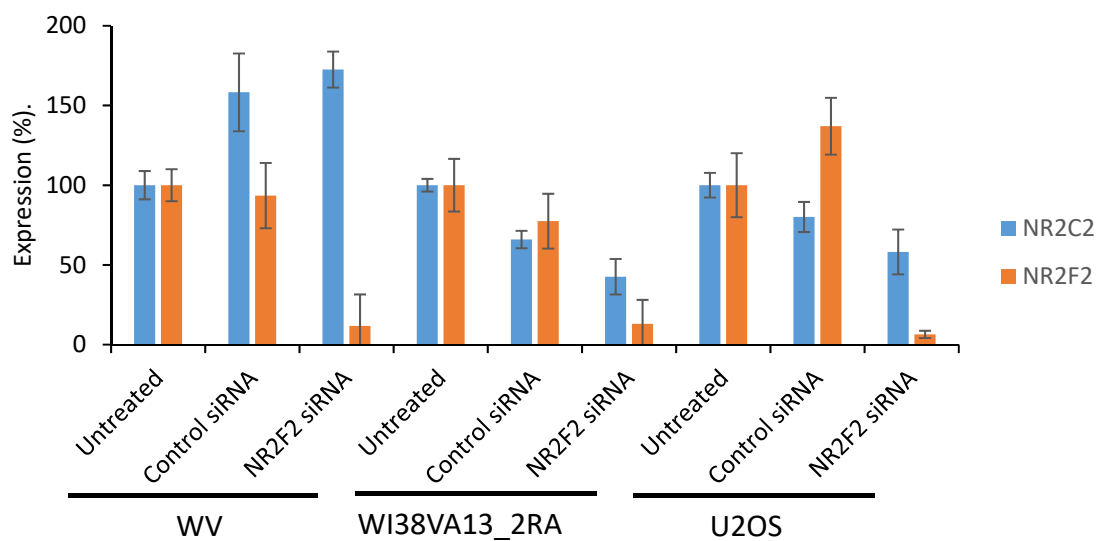
Figure 4.4 Quantification of siRNA-mediated NR2F2 depletion in ALT+ cell lines. (A) NR2C2 /NR2F2 western blot analysis of untreated, control siRNA, and NR2F2 siRNA treated cells. NR2C2 /NR2F2 protein expression was quantified using GAPDH as a loading control, (B) The bar charts show the protein expression as a percentage of the

expression level in untreated cells. Standard error bars are shown for three biological replicates.

4.3.3 C-circles analysis of siNR2F2 in ALT+ cell lines

In order to understand the relationship between NR2F2 expression and ALT activity, various ALT markers were measured after transient downregulation of NR2F2.

The C-circle assay shown in Figure 4.5, was carried out as described in section 2.2.2.11 on digested genomic DNA from siNR2F2 treated and untreated WV, WI38VA13_2RA, U2OS, and HT1080 cells. Triplicate dot blot assays showed a significant increase in relative abundance of c-circles abundance in U2OS ($p<0.0001$) and to a lesser extent in WV ($p<0.001$) compared to the control siRNA (Figure 4.6). This could suggest that NR2F2 plays a role in the regulation of some ALT related functions in these two cell lines. In contrast, the lower level of C-circles detected in the WI38VA13_2RA cell line did not change significantly (even when dCTP was added to the reaction), consistent with what has been reported in this particular cell line Conomos *et at.* (2012).

C-circle assay

Φ + dATP, dGTP+dTTP+ dCTP

3' Φ

C-circle → C-circle

5'

No primer (auto-priming)

$(TTAGGG)_n$ Telomeric G strand
$(CCCTAA)_n$ Telomeric C strand
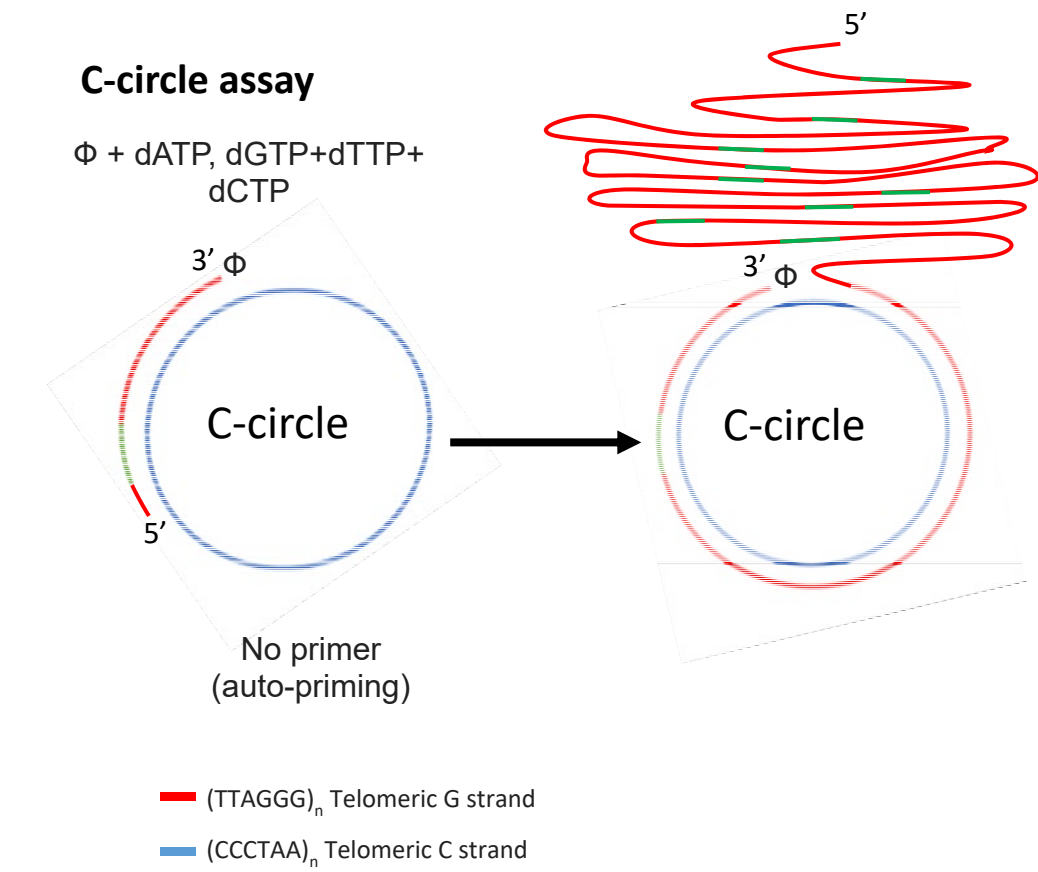
Figure 4.5 Schematic representation of the C-circle assay. The phi29 DNA polymerase (Φ) is used to amplify partially double-stranded C-circles by rolling circle replication. Green regions showed expected $(TCAGGG)_n$ or other variant repeats may be present in c-circles of the WI38VA13_2RA and possibly other ALT+ cell lines.
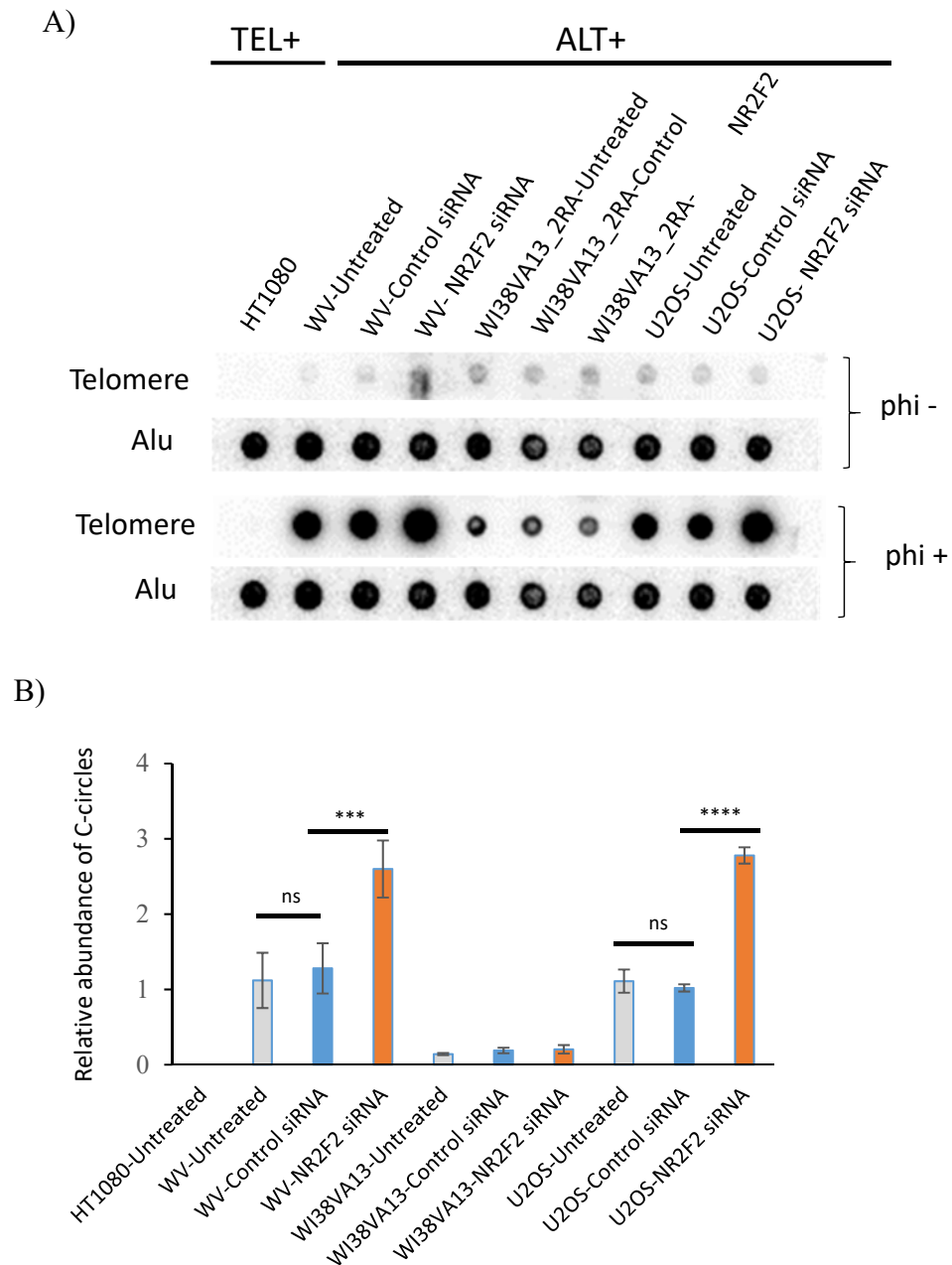
Figure 4.6 Quantification of the relative abundance of C-circles upon NR2F2 downregulation in ALT+ cell lines. (A) Example of a dot blot detection used to detect the rolling circle amplification products (with/without phi29 enzyme). Telomere, native dot blot hybridised to the telomere probe. Alu, denatured dot blot hybridised to the Alu probe. HT1080 was used as a negative control, (B) Bar chart representation of the mean of normalised C-circles levels ± standard deviation, n = 3 independent experiments. Two-way ANOVA tests were used to calculate p values. ns= not significant, ***=p<0.001, ****=p<0.0001.

4.3.4 Analysis of APBs in siNR2F2 treated ALT+ cell lines

The APB assay was used to detect a possible change in ALT activity following transient downregulation of NR2F2 in ALT+ cell lines. The frequency of nuclei with APBs was measured in cells seeded onto coverslips and then left untreated, or treated with the control siRNA or NR2F2 siRNA for 72 hours (Figure 4.7). Furthermore, confocal imaging using z-stack was used to verify the colocalisation of PML protein with telomeric DNA in APBs in some cells. This analysis showed small PML bodies with no telomeric DNA in HT1080, but large APBs containing PML protein coating a large mass of telomeric DNA in U2OS (Figure 4.8).
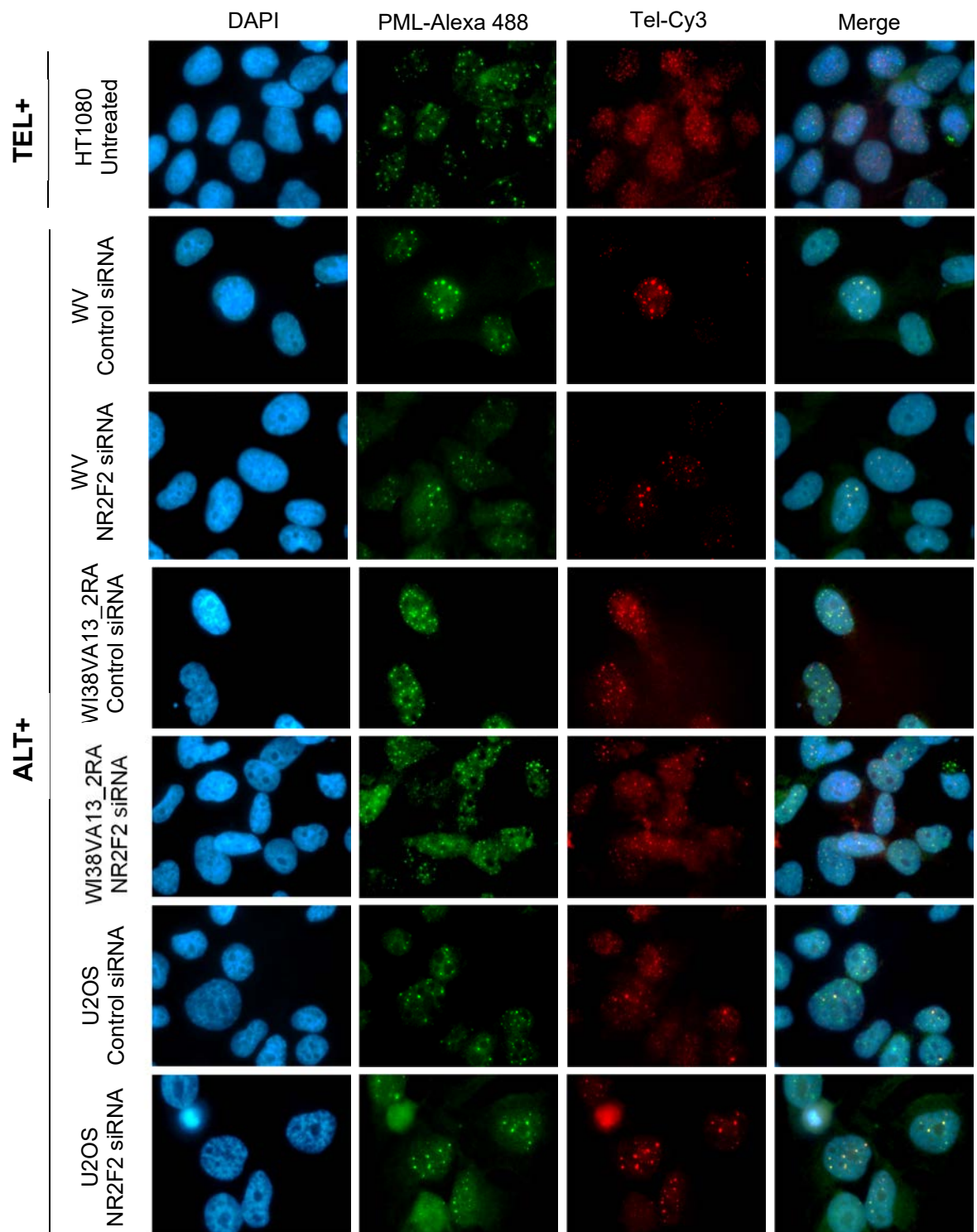
Figure 4.7 Detection of APBs upon NR2F2 downregulation in ALT+ cell lines.

IF staining of the PML protein (green, Alexa 488) with detection of telomeric DNA using a Telo-PNA probe (Cy3–(CCCTAA)$_3$, red) at interphase, and counterstained with DAPI (blue).
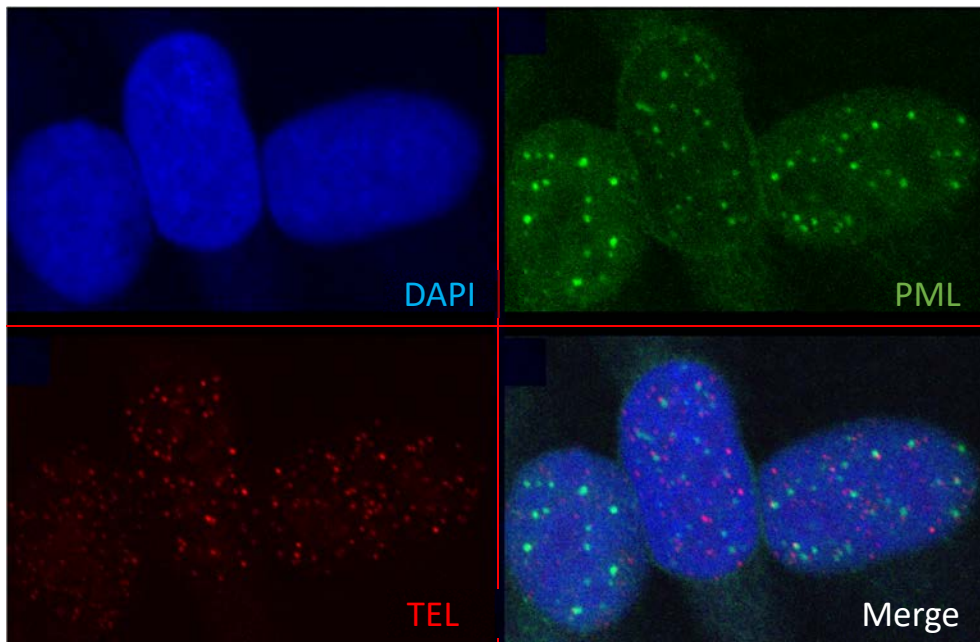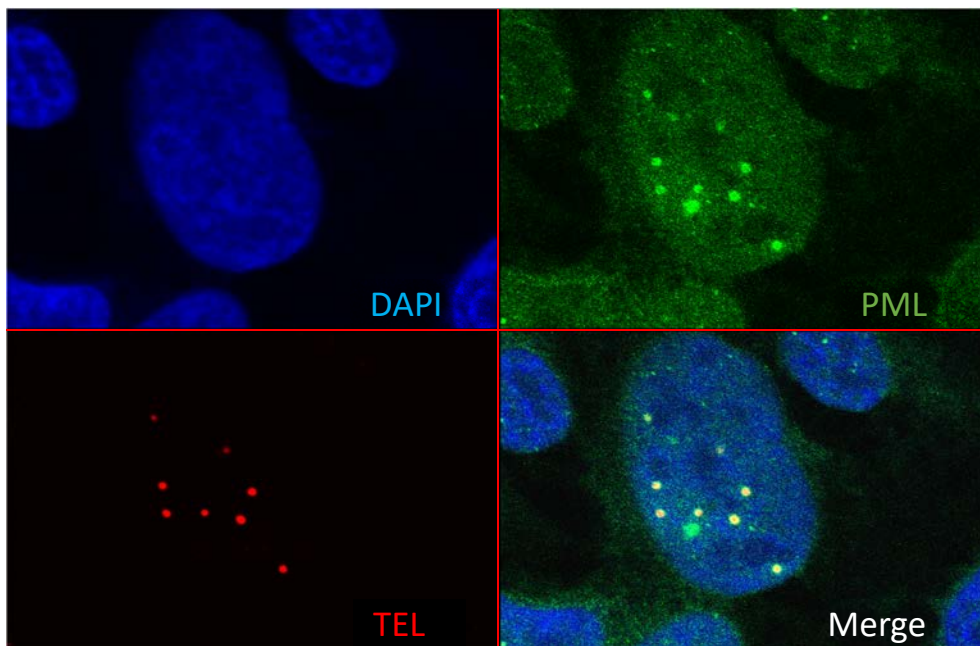
Figure 4.8 Visual inspection of APBs using confocal imaging. (A) HT1080 Tel+ cells, (B) U2OS ALT+ cells. A single slice from a focal stack with resolution of 0.255μm was captured. Blue channel (DAPI staining), green channel (IF staining of PML protein with Alexa 488), red channel (Telomere staining with Cy3-Telo-PNA).

As describe before, the changes to APBs were assessed measuring the average number of APBs per nucleus and the percentage of nuclei with APBs. The results are summarised in Table 4.1. A significant reduction in number of APBs per nucleus ($p<0.001$), and in frequency of nuclei with APBs (from ~9% to ~4%) was observed in WI38VA13_2RA ($p<0.01$), as reported previously (Conomos *et al.*, 2012). In contrast, no significant change in number of APBs per nucleus, nor in the frequency of nuclei with APBs were detected in the WV cells whereas a significant increase in the number of APBs per nucleus ($p<0.01$), and a large increase (~ 4-fold change; from ~7% to ~34%; ($p<0.0001$)) in frequency of nuclei with APBs was observed in U2OS cells (Figure 4.9).

Table 4.1 Summary of the APBs analysis after transient downregulation of NR2F2 in ALT+ cell lines.

Cell line	Condition	No of biological replicates	Total number of nuclei screened	No of APBs per nucleus containing APBs	Frequency of nuclei with APBs (%)
WV	Untreated	2	17068	8.50	17.10
WV	Control siRNA	3	12873	8.80	15.71
WV	NR2F2 siRNA	3	6288	9.63	16.25
WI38VA13_2RA	Untreated	2	8490	18.47	9.83
WI38VA13_2RA	Control siRNA	3	6025	18.49	8.76
WI38VA13_2RA	NR2F2 siRNA	3	11008	13.44	3.72
U2OS	Untreated	2	17817	7.10	7.06
U2OS	Control siRNA	3	18933	8.47	6.66
U2OS	NR2F2 siRNA	3	13248	11.69	34.24

A)

B)
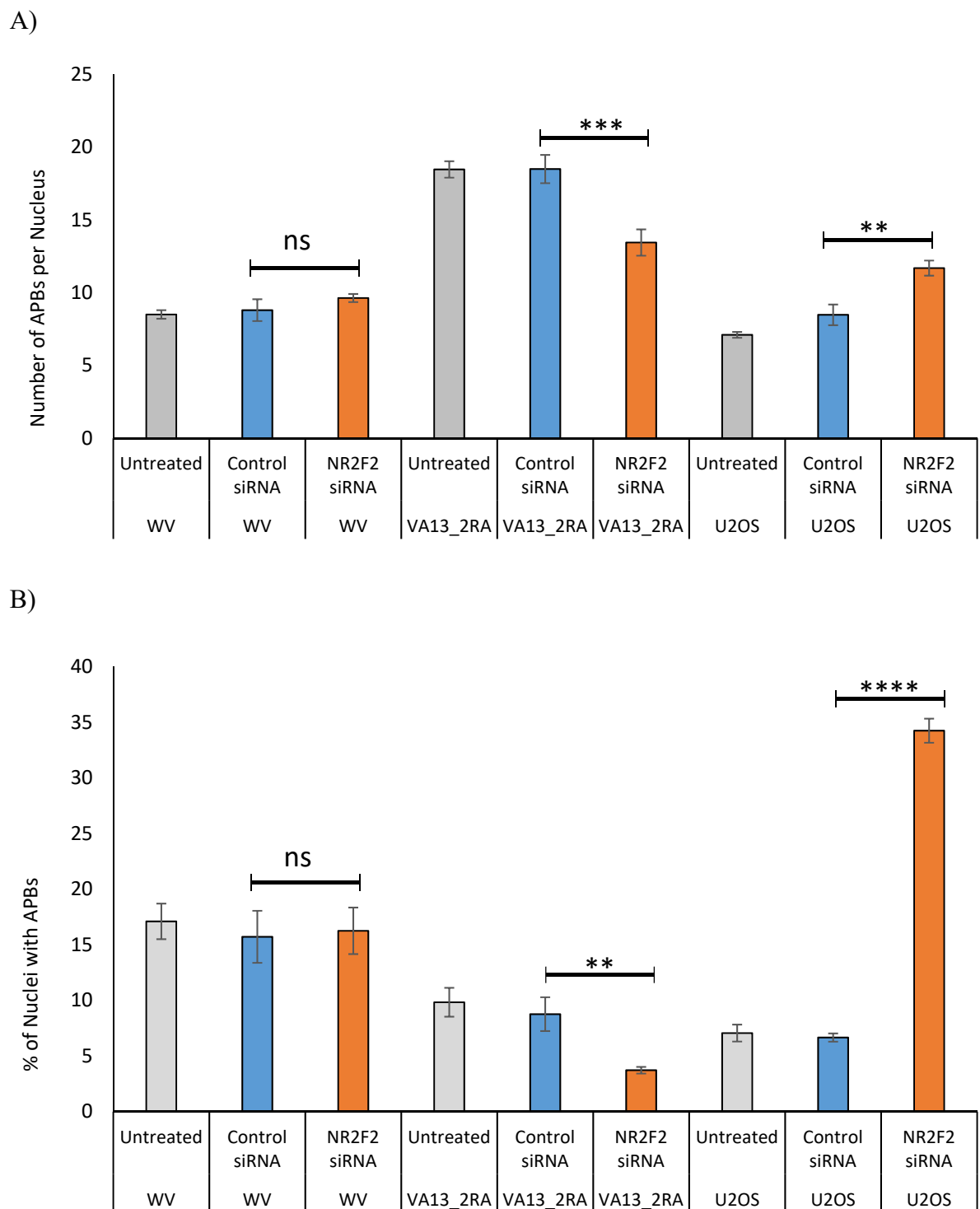
Figure 4.9 Quantification of APBs upon NR2F2 depletion in ALT+ cell lines. (A) The number of APBs per nucleus is shown as mean ± SD, (B) Frequencies of nuclei with APBs is shown as a percentage ± SD. Two-way ANOVA test were used to calculate p values. Control siRNA assays were used as reference group *=p<0.05, **=p<0.01, ***=p<0.001, ****=p<0.00001.

4.3.5 Cell cycle analysis in ALT+ cell lines following siNR2F2 treatment

It has been shown that NR2F2 promotes cell proliferation in cancer cells by modulating the expression of cell cycle regulator E2F1 and other genes (Johnson et al., 1993), cell cycle analysis was performed after 72 hours of siNR2F2 treatment. Cells were fixed in ethanol and stained with propidium iodide prior to FACS analysis as described in section 2.2.1.8.

The siNR2F2 treated WV cells showed a ~5% reduction in cells in S phase (p<0.0001), a ~ 4.8% increase in the number dead cells but no significant change in the percentage of cells in G1or G2 compared to control siRNA treated cells. The WI38VA13_2RA siNR2F2 treated cells showed a ~5.2% reduction in cells in S phase, a ~ 4.6% reduction in cells in G2 (p<0.0001) and a ~9.2% significant increase in cells in G1 (p<0.0001). In U2OS siNR2F2 treated cells, the changes to the cell cycle profile were more distinct. It showed a ~ 13% significant decrease in U2OS cells in S phase, an increase by ~ 9% and ~ 4% in cells in G1 and G2 respectively and no increase in percentage of dead cells (Figure 4.10, Figure 4.11, and Table 4.2).

The decrease in the percentage of cells with S phase in all three cell lines has suggested reduction in proliferation signalling as a result of depletion of NR2F2. This was consistent with NR2F2 role in regulating the cell cycle especially the genes that responsible for G1/S transition (Johnson *et al.*, 1993). However, the most remarkable effect and cell cycle arrest in G2 in U2OS cells suggested more extended role for NR2F2 in these cells. It was found that the cell cycle arrest in G2 in ALT+ cells is associated with depletion of the single-stranded DNA-binding protein replication protein A (RPA) (Grudic *et al.*, 2007). These possibilities would be investigated further by analysis of differential gene expression.
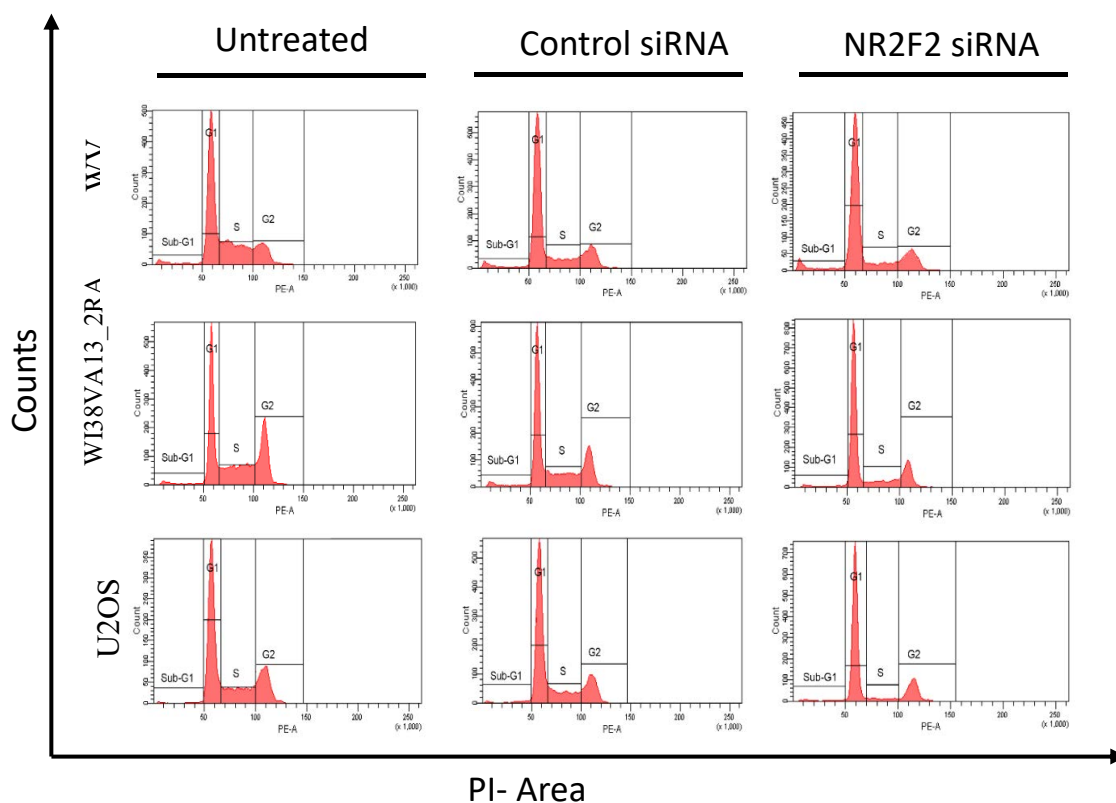
Figure 4.10 Cell cycle profiles of three ALT+ cells line treated with the siNR2F2. The data shows the density plots of cell cycle distribution after 72h of transient downregulation of NR2F2.

Table 4.2 Percentage of cells in phases of the cell cycle following no treatment, or treatment with control siRNA or siNR2F2. Data represented as mean ± standard deviation, n = 3 independent experiments.

Cell line	Condition	Sub-G1 (%)	G1 (%)	S (%)	G2 (%)
WV	Untreated	3.5±0.52	51.61±0.88	28.88±0.85	15.17±0.38
	Control siRNA	4.49±0.81	58.99±0.84	17.02±0.29	19.13±0.62
	NR2F2 siRNA	9.29±0.59	60.92±0.57	12.94±0.52	16.51±0.67
WI38VA13_2RA	Untreated	3.12±0.4	39.34±0.27	29.4±0.66	28.32±0.64
	Control siRNA	3.99±0.28	53.03±0.45	19.74±0.29	22.97±0.79
	NR2F2 siRNA	4.61±0.63	62.22±0.98	14.49±0.6	18.41±0.22
U2OS	Untreated	1.53±0.24	56.3±0.2	21.18±0.48	20.42±0.48
	Control siRNA	3.16±0.39	55.36±0.8	19.52±0.32	21.57±0.58
	NR2F2 siRNA	2.32±0.61	64.18±0.89	6.58±0.28	26.71±0.48
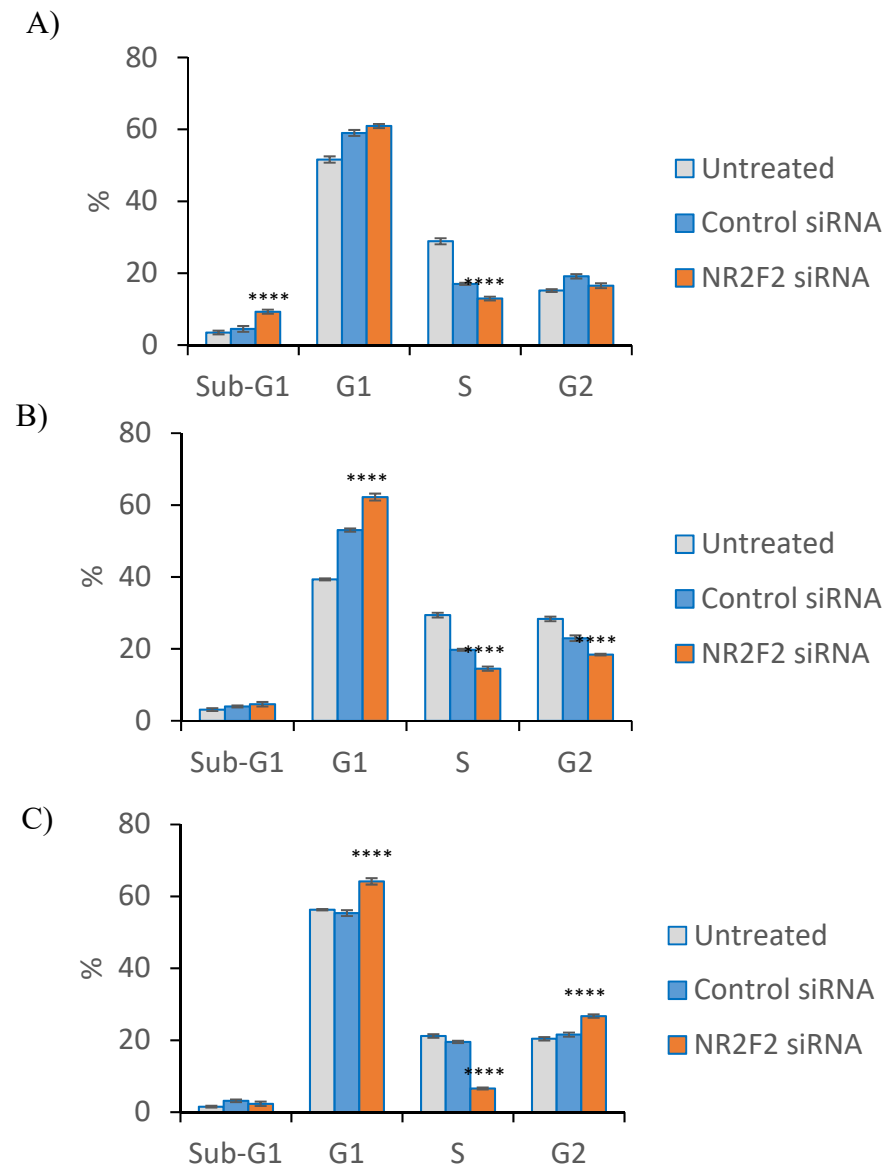
Figure 4.11 Cell cycle analysis of in ALT+ cell lines with NR2F2 depletion. (A) Cell cycle analysis of the WV cell line, (B) the WI38VA13_2RA cell line, (C) the U2OS cell line. Bar chart mean ± standard deviation, n = 3 independent experiments. Two-way ANOVA test were used to calculate p values. Control siRNA as reference group *=p<0.05, **=p<0.01, ****=p<0.00001.

4.3.6 Gene expression profiling in NR2F2 depleted ALT+ cell lines

4.3.6.1 Alignment summary of RNA-seq

To obtain robust data from assessment of changes in gene expression three biological replicates are required. Therefore, three biological replicates were send to Novogene for non-stranded mRNA sequencing as described in section 2.2.3.2. For each sample there were over 30 million paired-end (PE) reads, which passed the Q20 filter, of which 82-89% were uniquely mapped to human genome reference, 1.4-2.4% of the reads mapped to multiple loci and 9-14% of the reads did not map to any region in the human genome (Appendix 4.2). The uniquely mapped reads were equally distributed to forward and reverse strands in all the samples. The comparative transcription analysis was conducted by counting the reads that mapped uniquely to genes, which was 80-86% of all reads.

4.3.6.2 Sample-sample correlation analysis

After alignment and QC analyses of the RNA-seq data, a gene counts matrix was generated of uniquely mapped PE reads for gene expression quantification and deferential expression analysis. The Fragments Per Kilobase of transcript per Million mapped reads (FPKM) method was used for gene expression quantification, as the number of reads counted were normalised to gene length. FPKM normalised gene expression was used for sample-sample correlation analysis to assess variation between the biological replicates and between samples, see (Appendix 4.3).

The correlation coefficient was greater than 0.96 between samples of each group, while the dissimilarity (represented by R^2) between the Control siRNA replicates and NR2F2 siRNA replicates was < 0.01. This was an indicator for the high repeatability and reliability of this experiment. It also showed that WV is closer in expression profile to WI38VA13/2RA than U2OS (Appendix 4.4).These relations were also visualised by dendrogram clusters and 2D principal components plot (Figure 4.12). The fact that both WV and WI38VA13 are SV40-trasformed fibroblast cell lines, while U2OS is derived from osteosarcoma could explain these findings. However, the effect of siRNA NR2F2 on gene expression was more predicated in U2OS than the other two ALT+ cell lines.

A)
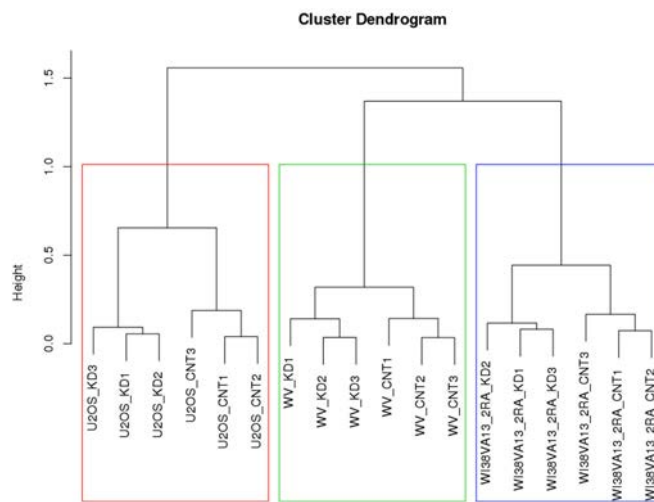
B)
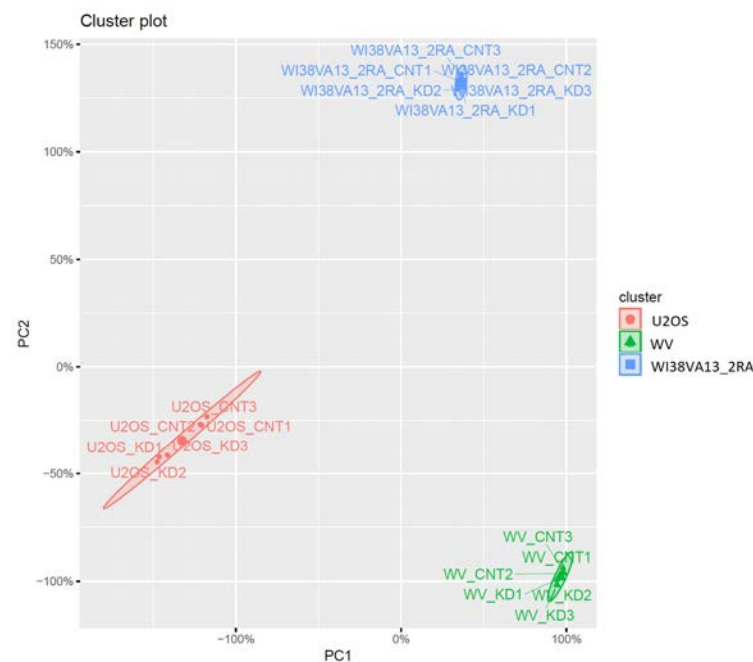
Figure 4.12 Sample-to-sample distance analysis of RNA-seq data from siNR2F2 in ALT+ cell lines. (A) Dendrogram clusters of RNA-seq data from all the samples. Sample-to-sample distance has calculated from log2 of normalised read count matrix. The hierarchical clustering with complete linkage method was used to plot the dendrogram, (B) Principal component analysis of log2 of normalised read count matrix.

It shows the distance between the samples in the 2D plane based on their first two principal components.

4.3.6.3 Differential gene expression analysis following NR2F2 depletion

The differential expression analysis between control siRNA treated and siNR2F2 treated cells was conducted as described in 2.2.3.2 using DESeq2 tool. The false discovery rate (FDR) less than 0.05 was used as the threshold. To measure the effect of siNR2F2 treatment on gene expression, the frequency of differently expressed genes were calculated using different fold changes as cut-off points (log2 fold change >0, >1, >2). The summaries for differentially expressed genes is shown in Table 4.3. Overall, similar frequencies of significantly down- and up-regulated genes following NR2F2 depletion were observed in all three cell lines. At log2 fold change>0, ~42%, 52%, and 61% of expressed genes were differently expressed in WV, WI38VA13, and U2OS cells respectively. However, much variable frequencies of differently expressed genes were observed at cut-off point log2 fold change > 1 (more than 2 fold changes), about 1.5%, 7%, and 21% respectively in WV, WI38VA13, and U2OS. This, in conjunction with the observed cell cycle arrest at G1 and G2 phases only in the U2OS cell line, suggests that transient NR2F2 downregulation had a greater effect in U2OS than in WV or WI38VA13_2RA.

Table 4.3 Summary of DE genes upon NR2F2 downregulation in ALT+ cell lines. It showed the number and percentage of DE genes as a proportion of the total number of expressed genes at different cut-off points of log2 fold change and FDR <0.05

	# Expressed Genes	Status	DE genes, log2 fold change					
			>0		>1		>2	
			#	%	#	%	#	%
WV	**13376**	Down	2681	20.04	41	0.31	0	0.00
		UP	2868	21.44	152	1.14	3	0.02
		Total	**5549**	**41.48**	**193**	**1.45**	**3**	**0.02**
WI38VA132_2RA	**14087**	Down	3555	25.24	330	2.34	23	0.16
		UP	3725	26.44	601	4.27	82	0.58
		Total	**7280**	**51.68**	**931**	**6.61**	**105**	**0.75**
U2OS	**16303**	Down	5049	30.97	1765	10.83	408	2.50
		UP	4915	30.15	1707	10.47	444	2.72
		Total	**9964**	**61.12**	**3472**	**21.30**	**852**	**5.23**

To identify the genes that have been effected by downregulation of NR2F2 in these three cell lines, the intersection of the differentially expressed genes at log2 fold change>1 in the three cell lines was carried out. This showed that only 72 genes (~ 2% of total DE genes) are in common (Figure 4.13 and Table 4.4). The manual classification of these 72 genes showed that ten are involved in transmembrane activity and cell-cell adhesion (*TMEM87B, TMEM87B, TMEM109, SLC16A2 , SLC17A5, CRIM1, LPP, CACNA1G, NECTIN4, GPR155*); three are involved in metabolism (*ACLY, IP6K3, ABCA12*); eight are involved in protein metabolism (*FBXO33, ADAMTSL1, SCG5, FBXO32, USP18, PPP1R2, BRI3BP, MAPK6*); four are involved in cytokine and inflammatory response (*GBP1, SP110, IL1R1, TRIM22*); three interact with tubulin and actin (*TUBE1, CGN, ACTA2*); three are present in synaptic vesicles (*SYP, SYNGR3, ABAT*); five genes are shown to be involved in regulation of gene expression (*UTP18,COMMD8, NCOA7, BTG2, EIF1AX*); four genes are involved in cell development and differentiation (*MYADM, CD2AP, DACT3, PTCHD4*); three genes encodes for growth factors (*DAZAP2, CSF1, TGFBR1*); five genes regulate cell death (*PDCD4, INPP5D, CYFIP2,TP53INP1, FAS*); and three genes are involved in cell cycle (*NEK7, CDKN1A/p21, REC8*). There was also a change in the expression of the UNG gene that plays a role in C-strand Synthesis of telomeres.
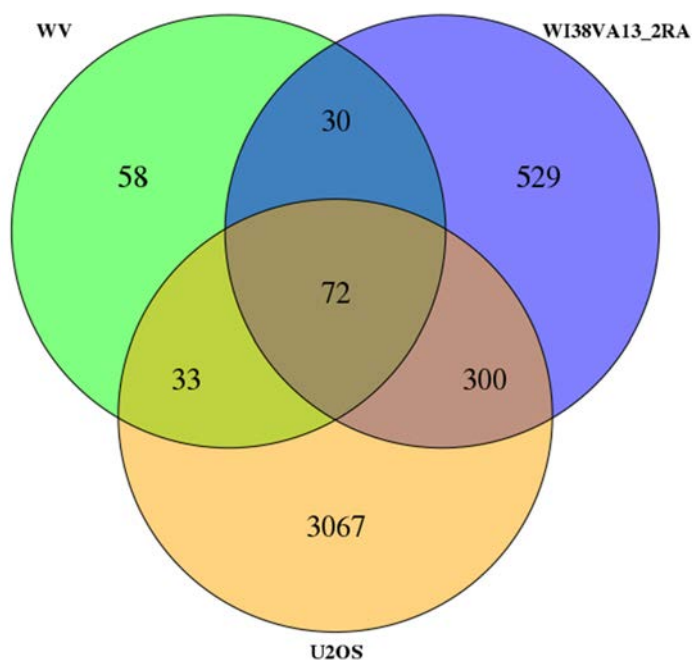
Figure 4.13 Intersection of DE genes share between the ALT+ cell lines following NR2F2 depletion. Venn diagram showing the overlap of genes that showed at log2 fold change >1 change in gene expression (FDR <0.05).

Table 4.4 List of the 72 DE genes shared by all three ALT+ cell lines upon NR2F2 depletion. Threshold, log2 fold change > 1, FDR <0.05.

Ensembl Gene ID	Gene symbol	Description
ENSG00000147100	SLC16A2	solute carrier family 16 member 2
ENSG00000131473	ACLY	ATP citrate lyase
ENSG00000179820	MYADM	myeloid associated differentiation marker
ENSG00000076248	UNG	uracil DNA glycosylase
ENSG00000035141	FAM136A	family with sequence similarity 136 member A
ENSG00000183283	DAZAP2	DAZ associated protein 2
ENSG00000185551	NR2F2	nuclear receptor subfamily 2 group F member 2
ENSG00000153214	TMEM87B	transmembrane protein 87B
ENSG00000011260	UTP18	UTP18, small subunit processome component
ENSG00000164061	BSN	bassoon presynaptic cytomatrix protein
ENSG00000151414	NEK7	NIMA related kinase 7
ENSG00000124762	CDKN1A	cyclin dependent kinase inhibitor 1A
ENSG00000180694	TMEM64	transmembrane protein 64
ENSG00000119899	SLC17A5	solute carrier family 17 member 5
ENSG00000150938	CRIM1	cysteine rich transmembrane BMP regulator 1
ENSG00000150593	PDCD4	programmed cell death 4
ENSG00000174574	AKIRIN1	akirin 1 - unknown function
ENSG00000110108	TMEM109	transmembrane protein 109
ENSG00000165355	FBXO33	F-box protein 33
ENSG00000198087	CD2AP	CD2 associated protein
ENSG00000164983	TMEM65	transmembrane protein 65
ENSG00000156642	NPTN	neuroplastin
ENSG00000184992	BRI3BP	BRI3 binding protein
ENSG00000184203	PPP1R2	protein phosphatase 1 regulatory inhibitor subunit 2
ENSG00000132718	SYT11	synaptotagmin 11
ENSG00000106799	TGFBR1	transforming growth factor beta receptor 1
ENSG00000128965	CHAC1	ChaC glutathione specific gamma-glutamylcyclotransferase 1
ENSG00000159339	PADI4	peptidyl arginine deiminase 4
ENSG00000132274	TRIM22	tripartite motif containing 22
ENSG00000173905	GOLIM4	golgi integral membrane protein 4
ENSG00000173674	EIF1AX	eukaryotic translation initiation factor 1A, X-linked
ENSG00000069956	MAPK6	mitogen-activated protein kinase 6
ENSG00000161896	IP6K3	inositol hexakisphosphate kinase 3

ENSG00000259529	N/A	Uncharacterised gene
ENSG00000168918	INPP5D	inositol polyphosphate-5-phosphatase D
ENSG00000178031	ADAMTSL1	ADAMTS like 1
ENSG00000105270	CLIP3	CAP-Gly domain containing linker protein 3
ENSG00000024526	DEPDC1	DEP domain containing 1
ENSG00000169019	COMMD8	COMM domain containing 8
ENSG00000145012	LPP	LIM domain containing preferred translocation partner in lipoma
ENSG00000074935	TUBE1	tubulin epsilon 1
ENSG00000117228	GBP1	guanylate binding protein 1
ENSG00000135899	SP110	SP110 nuclear body protein
ENSG00000143375	CGN	cingulin
ENSG00000048392	RRM2B	ribonucleotide reductase regulatory TP53 inducible subunit M2B
ENSG00000197380	DACT3	dishevelled binding antagonist of beta catenin 3
ENSG00000111912	NCOA7	nuclear receptor coactivator 7
ENSG00000055163	CYFIP2	cytoplasmic FMR1 interacting protein 2
ENSG00000166922	SCG5	secretogranin V
ENSG00000164938	TP53INP1	tumor protein p53 inducible nuclear protein 1
ENSG00000102003	SYP	synaptophysin
ENSG00000159388	BTG2	BTG anti-proliferation factor 2
ENSG00000156804	FBXO32	F-box protein 32
ENSG00000100918	REC8	REC8 meiotic recombination protein
ENSG00000130518	KIAA1683	KIAA1683 – unknown function
ENSG00000006283	CACNA1G	calcium voltage-gated channel subunit alpha1 G
ENSG00000143217	NECTIN4	nectin cell adhesion molecule 4
ENSG00000163328	GPR155	G protein-coupled receptor 155
ENSG00000188610	FAM72B	family with sequence similarity 72 member B – unknown function
ENSG00000144452	ABCA12	ATP binding cassette subfamily A member 12
ENSG00000026103	FAS	Fas cell surface death receptor
ENSG00000107796	ACTA2	actin, alpha 2, smooth muscle, aorta
ENSG00000127561	SYNGR3	synaptogyrin 3
ENSG00000280744	LINC01173	long intergenic non-protein coding RNA 1173
ENSG00000196550	FAM72A	family with sequence similarity 72 member A – unknown function
ENSG00000183044	ABAT	4-aminobutyrate aminotransferase
ENSG00000244694	PTCHD4	patched domain containing 4
ENSG00000115594	IL1R1	interleukin 1 receptor type 1
ENSG00000215784	FAM72D	family with sequence similarity 72 member D – unknown function
ENSG00000184979	USP18	ubiquitin specific peptidase 18
ENSG00000263513	FAM72C	family with sequence similarity 72 member C – unknown function
ENSG00000184371	CSF1	colony stimulating factor 1

4.3.6.4 Analysis of changes in the gene expression profile of WV cells

The data of DE genes are also shown in volcano plots (Figure 4.14). The genes that are labelled in the volcano plots are top 30 highly differently expressed (based on p value).
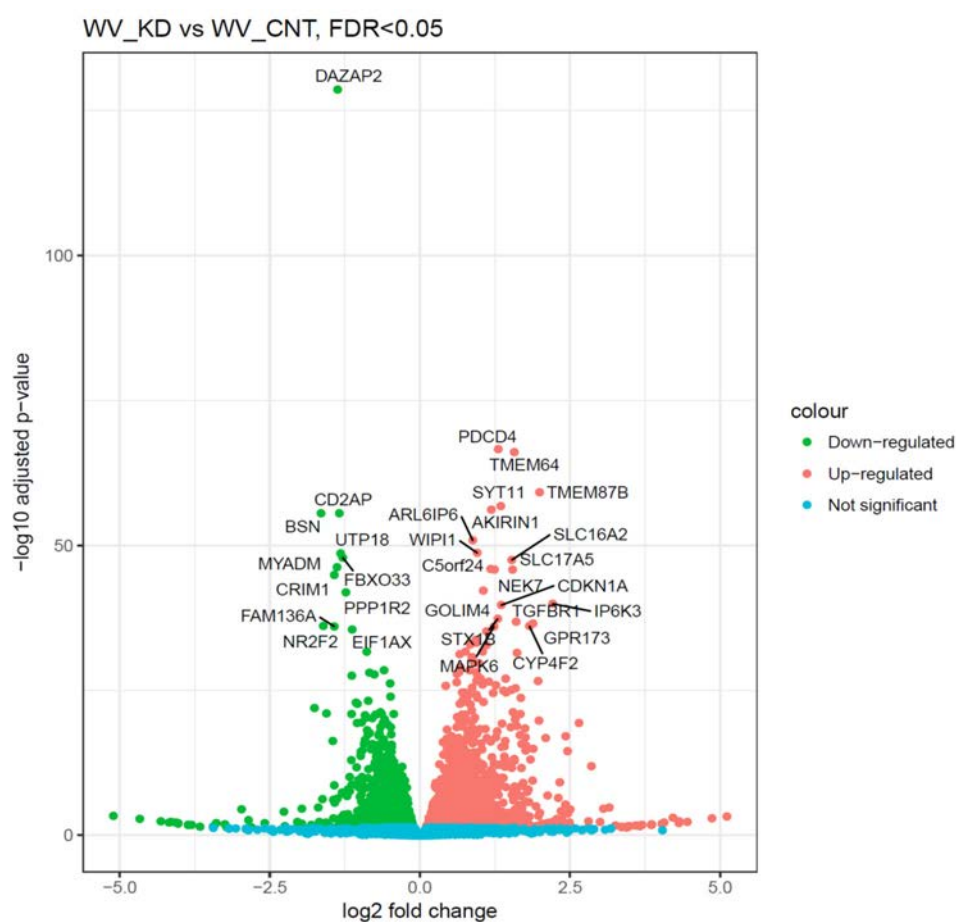
Figure 4.14 Volcanoplot of DE genes at depletion of NR2F2 in WV cell line.Top 30 DE genes are labelled.

To characterise the biological process that have changed as a result of NR2F2 depletion in WV cells, GO enrichment analysis was preformed using of DE genes (846 DE genes, fold change >1.6, FDR <0.05) identified change in expression of several biological processes (Figure 4.15, and Appendix 4.5). The results of this analysis showed changes in biological processes that mainly encompassed genes related to blood vessel morphogenesis (Downregulated: *NPPB, E2F8*; Upregulated: *VASH2, PDGFRA, EPHB2*), large group of protein kinases (downregulated: *CDC25A, CKS2, DUSP7, LPAR1, FAM72C, CDK5R2, SFN, FAM72C, PDGFD, NSUN5P1, PIK3CG*; upregulated: *SHC1, CDKN1A, PDCD4, TGFBR1, EIF2AK2, TNFRSF10B, MAP3K12, IQGAP1, DUSP10, FAM212B, TRIB2. ALK, FRS2, PTPN22, RGS3, HSPB1, IRAK2, CSF1R, DUSP8, MAPK6*), of development and differentiation (downregulated: *HES1, ODC1, WNK4, EIF2AK3, PEX7, CHSY1;* upregulated*; STX1B, RRM2B, ADAMTS1, EPHB2, LGR5, TP73, ACTA2, JAG1, SMAD5, PECAM1, GDPD2, TMEM64, CTHRC1, ITGA11*), extracellular matrix and cell migration **(**downregulated: *COL9A2, BCL3;* upregulated*: ITGA10, ITGA11, ITGAX, MMP8, PLFML2A***),** multicellular organism process **(**downregulated: *NR2F2, FOS, GCLM, ITGA2, HSF1, STAT5B.* upregulated*: CHRHBP***),** and to a lesser extent in negative regulation of G1/S transition of mitotic cell cycle (*CDKN1A, CDKN2B, FBXO7, CCNB1, MDM2*), and positive regulation of cell death (*PDCD4, FAS, DAPK1*). To reduce the number of genes that need to be analysed, the clusterprofiler (an R packages) was used to extract the over-represented genes (core genes; genes that are present in most significantly enriched biological process) in GO enrichment of biological processes. The core of enriched biological processes is shown in Figure 4.16.
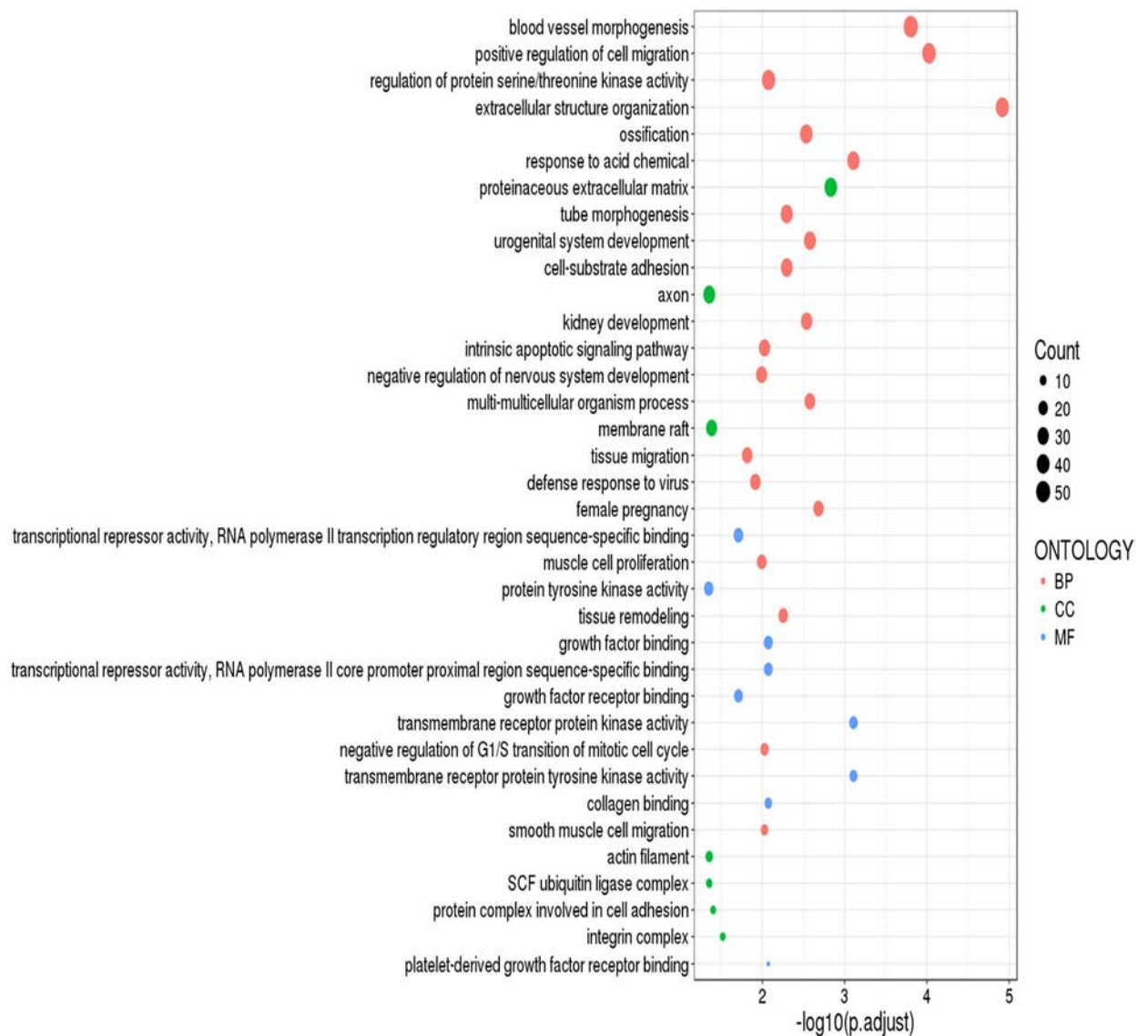
Figure 4.15 GO enrichment of DE genes following NR2F2 downregulation in WV. Differently expressed genes in the WV cell line (fold change >1.6, FDR <0.05) were mapped to GO terms using DAVID Bioinformatics Resources v6.8, where ALL GO terms were included in this analysis. The enrichment results were summarised by removing redundant GO terms (allowed similarity=0.7) using REViGO online tool (Supek *et al.*, 2011). GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.

Figure 4.16 Visualisation of DE genes in most significantly enriched biological processes that are associated with NR2F2 depletion in WV. It illustrates the biological complexities in which a gene may belong to multiple themes, it also shows the terms linked to the most significant biological processes associated with the genes that show DE upon downregulation of NR2F2 in WV cells. Fold change >1.6, FDR <0.05.

In order to investigate the relationship between the ALT related genes (73 genes), NR2F2-target genes (413 genes) and DE genes (fold change >1.6) in WV, triple intersection was carried out (Figure 4.17). The results showed no overlap between ALT related genes and DE genes. This suggest no direct role for NR2F2 in the regulation of ALT mechanism in WV. The intersection between the list of NR2F2 target genes and the DE genes identified only four genes (Table 4.5); two kinases (*MAPK6, TRIB2*), other two genes play role in exocytosis (*SIGLEC14*) and cell development (*STX1B*).
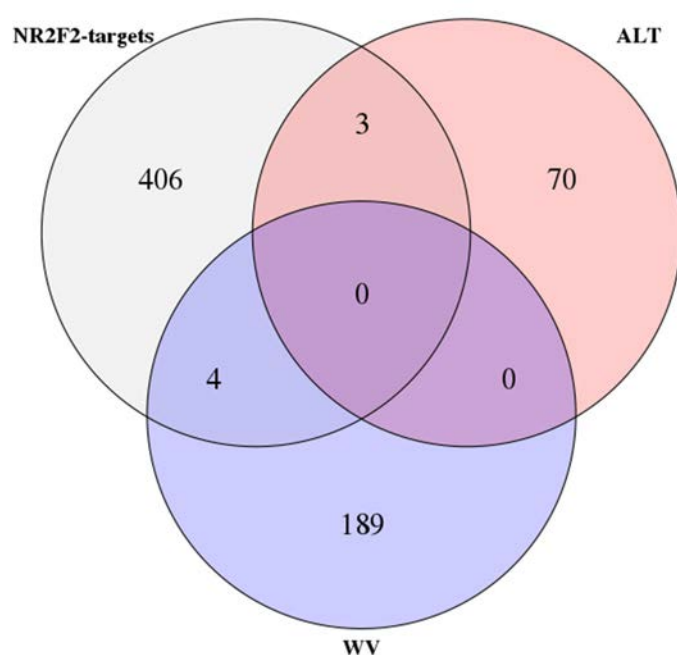
Figure 4.17 Intersection of DE genes in WV with ALT- related genes and NR2F2-trargets. Log2 fold change >1, FDR <0.05 .

Table 4.5 Intersection of NR2F2-target genes with DE genes in WV upon depletion of NR2F2.Log2 fold change >1, FDR <0.05.

Ensambl Gene ID	baseMean	log2FoldChange	p value	p adj	Gene symbol	Full Name
ENSG00000099365	637.94	1.31	1.46E-48	2.43E-46	STX1B	syntaxin 1B
ENSG00000069956	3532.52	1.25	3.14E-47	5.10E-45	MAPK6	mitogen-activated protein kinase 6
ENSG00000254415	139.64	1.89	1.08E-23	5.24E-22	SIGLEC14	sialic acid binding Ig like lectin 14
ENSG00000071575	343.52	1.12	1.97E-17	5.69E-16	TRIB2	tribbles pseudokinase 2

4.3.6.5 Analysis of changes in the gene expression profile of WI38VA13_2RA cells

The data of DE genes were also showed in volcano plots (Figure 4.18). The genes that are labelled in the volcano plot are top 80 highly differently expressed (based on p value).

Figure 4.18 Volcanoplot of DE genes at depletion of NR2F2 in WI38VA13_2RA cell line.Top 80 DE genes have labelled.

To characterise the biological process that have changed as a result of NR2F2 depletion in WI38VA13_2RA cells, GO enrichment analysis was preformed using the DE genes (931 DE genes, fold change >2, FDR <0.05) . This identified changes in several biological processes (Figure 4.19, and Appendix 4.6). The results of this analysis showed changes in biological processes that mainly encompassed genes related to blood vessel morphogenesis (downregulated; *NR2F2, EGF, PIK3CB, IL18, SEMA3E, TGFBR3, HOXA1, EIF2AK3, STRA6, TEK;* upregulated: *ITGB2, PLXDC1, NOTCH3, ESM1, SULF1/2, TGFBR1, HAS2*), regulation of cell migration and extracellular matrix organization (downregulated: *TIAM1, GCNT2, CAPN7, ROCK2, ANXA3, PDGFD, LYN, NOS1AP, CHMP4C, TNFAIP6, EPB41L4B, MYADM, MITF, TEK*; upregulated: *MMP9, TGFBR1, HAS2, CCL3, CPEB1, RET, FOXF1,CSF1, CXCL1,CCL5, ITFB2, EDNRB*), type I interferon signalling pathway (downregulated: *GBP2,* upregulated*: MX2, OAS1/2, IFIT1/2, IRF9, IFITM1*), cellular development and differentiation (downregulated: *EFEMP1, TULP3, HOXC10, HOXA1, EIF2AK3, ACVRL1, EFEMP1, DLL3, ZEB1, CHSY1;* upregulated*: HMGA2, MMP9, POSTN, ALX4*). To reduce the number of genes that need to be analysed, the clusterprofiler was used to extract the over-represented genes in GO enrichment of biological processes. The core of enriched biological processes is shown in Figure 4.20.
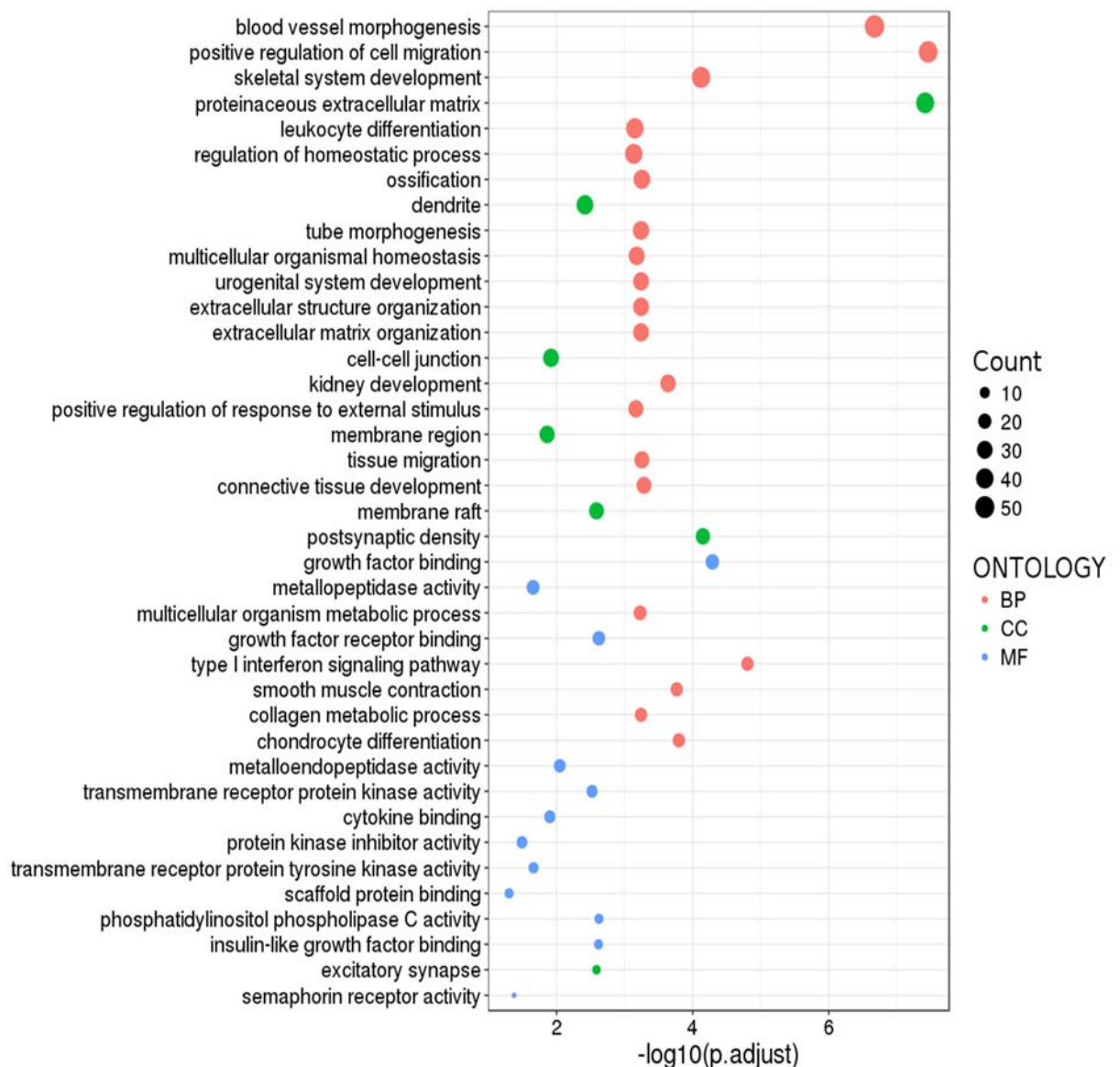
Figure 4.19 Analysis of GO enrichment terms associated with the DE genes upon NR2F2 downregulation in WI38VA13_2RA. Differently expressed genes in the WI38VA13_2RA cell line (fold change >2, FDR <0.05) were mapped to GO terms using DAVID. The enrichment results were summarised by allowed medium similarity=0.7 using REViGO online tool (Supek *et al.*, 2011). GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.
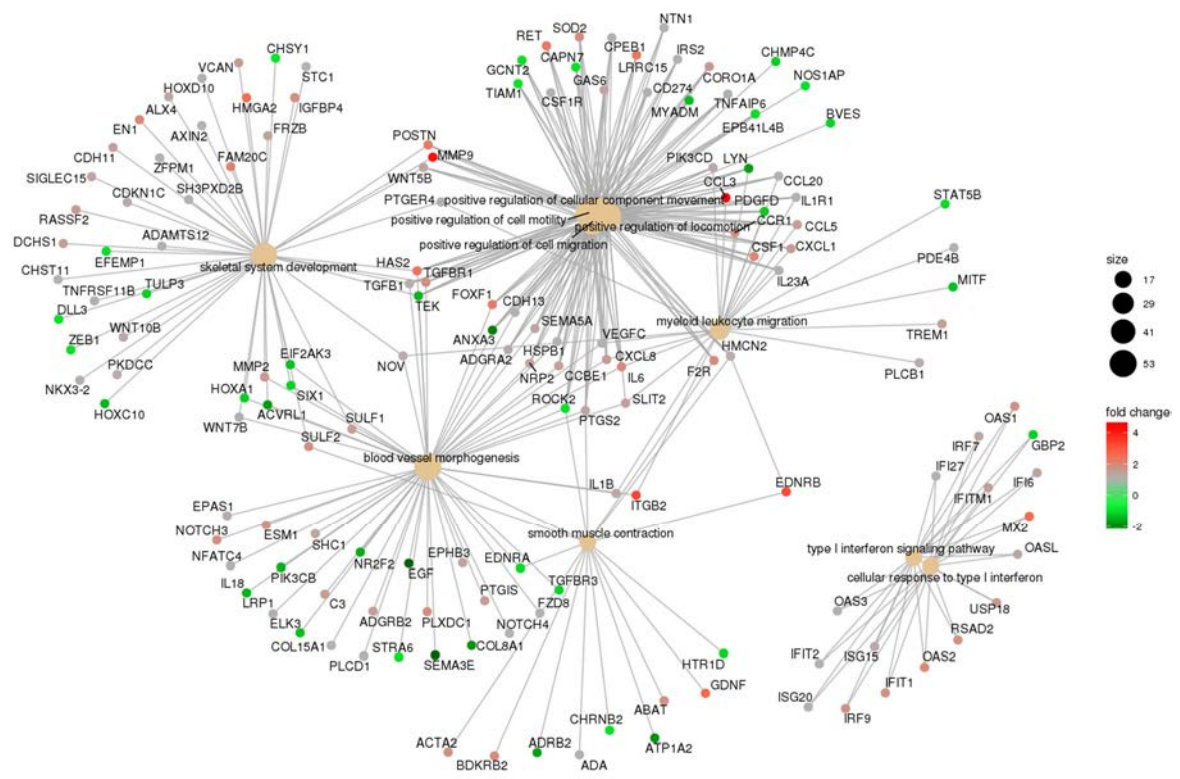
Figure 4.20 Visualisation of DE genes in most significantly enriched biological processes that associated with NR2F2 depletion in WI38VA13_2RA. Illustrates the biological complexities in which a gene may belong to multiple themes, it also shows the terms linked to the most significant biological processes associated with genes that show to DE upon downregulation of NR2F2 in WI38VA13_2RA cells. Fold change >2, FDR <0.05.

To investigate the relationship between the ALT related genes, NR2F2-target genes and DE genes (fold change >2) in WI38VA13_2RA a triple intersection was carried out (Figure 4.21). The results showed no overlap between the ALT related genes and DE genes. As for WV, this suggests no direct role for NR2F2 in regulation of ALT related genes in the WI38VA13_2RA cell line. However, the intersection between NR2F2 genes and DE genes identified 14 genes shown in (Table 4.6); two protein kinases (*MAPK6, TRIB2*), and other 12 genes that play roles in protein metabolism and cellular matrix and migration.
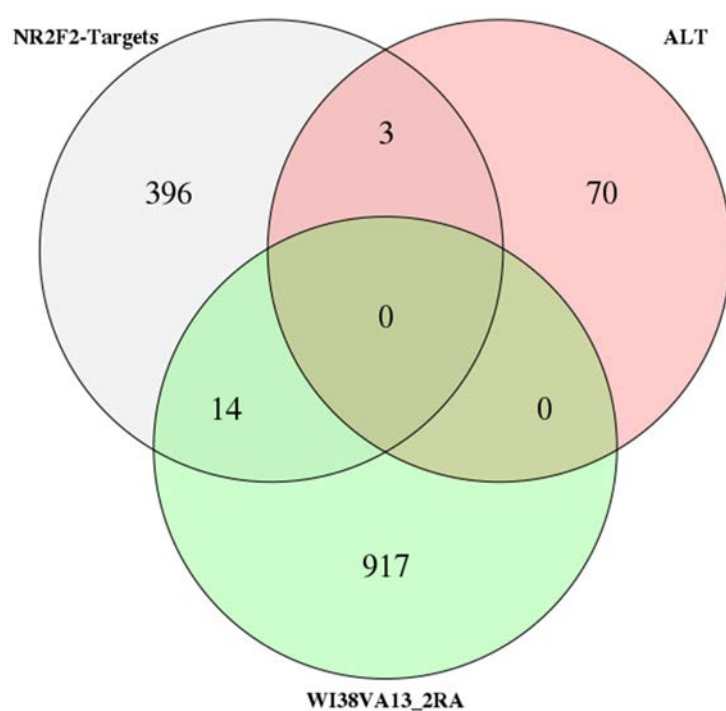
Figure 4.21 Intersection of DE genes in WI38VA13_2RA with ALT- related genes and NR2F2-trargets. Log2 fold change >1, FDR <0.05.

Table 4.6 Intersection of NR2F2 target genes with DE genes in WI38VA13_2RA upon depletion of NR2F2. Fold change >2, FDR <0.05.

Ensambl Gene ID	baseMean	log2FoldChange	p value	p adj	Gene symbol	Full Name
ENSG00000145147	4171.19	1.48	4.26E-83	1.43E-80	SLIT2	slit guidance ligand 2
ENSG00000069956	3488.03	1.07	1.32E-56	1.75E-54	MAPK6	mitogen-activated protein kinase 6
ENSG00000065308	6432.85	-1.05	9.74E-56	1.24E-53	TRAM2	translocation associated membrane protein 2
ENSG00000186815	2222.94	1.04	4.23E-46	3.51E-44	TPCN1	two pore segment channel 1
ENSG00000054356	163.59	3.96	9.78E-46	7.91E-44	PTPRN	protein tyrosine phosphatase, receptor type N
ENSG00000136859	758.13	1.30	3.18E-25	9.38E-24	ANGPTL2	angiopoietin like 2
ENSG00000184678	333.88	-1.33	3.37E-24	9.24E-23	HIST2H2BE	histone cluster 2 H2B family member e
ENSG00000104976	1259.29	1.11	5.37E-22	1.22E-20	SNAPC2	small nuclear RNA activating complex polypeptide 2
ENSG00000101230	193.61	1.54	2.54E-21	5.60E-20	ISM1	isthmin 1
ENSG00000235750	96.36	-1.73	2.63E-15	3.61E-14	KIAA0040	KIAA0040
ENSG00000178726	103.50	-1.46	1.14E-12	1.19E-11	THBD	thrombomodulin
ENSG00000071575	107.45	1.43	3.46E-10	2.73E-09	TRIB2	tribbles pseudokinase 2
ENSG00000099953	61.66	1.07	9.87E-05	3.42E-04	MMP11	matrix metallopeptidase 11
ENSG00000197046	31.37	1.32	1.77E-04	5.83E-04	SIGLEC15	sialic acid binding Ig like lectin 15

4.3.6.6 Analysis of changes in the gene expression profile of U2OS cells

The data for the DE genes are shown in volcano plot (Figure 4.22). The genes that are labelled in the volcano plot are top 30 highly differently expressed (based on p value).
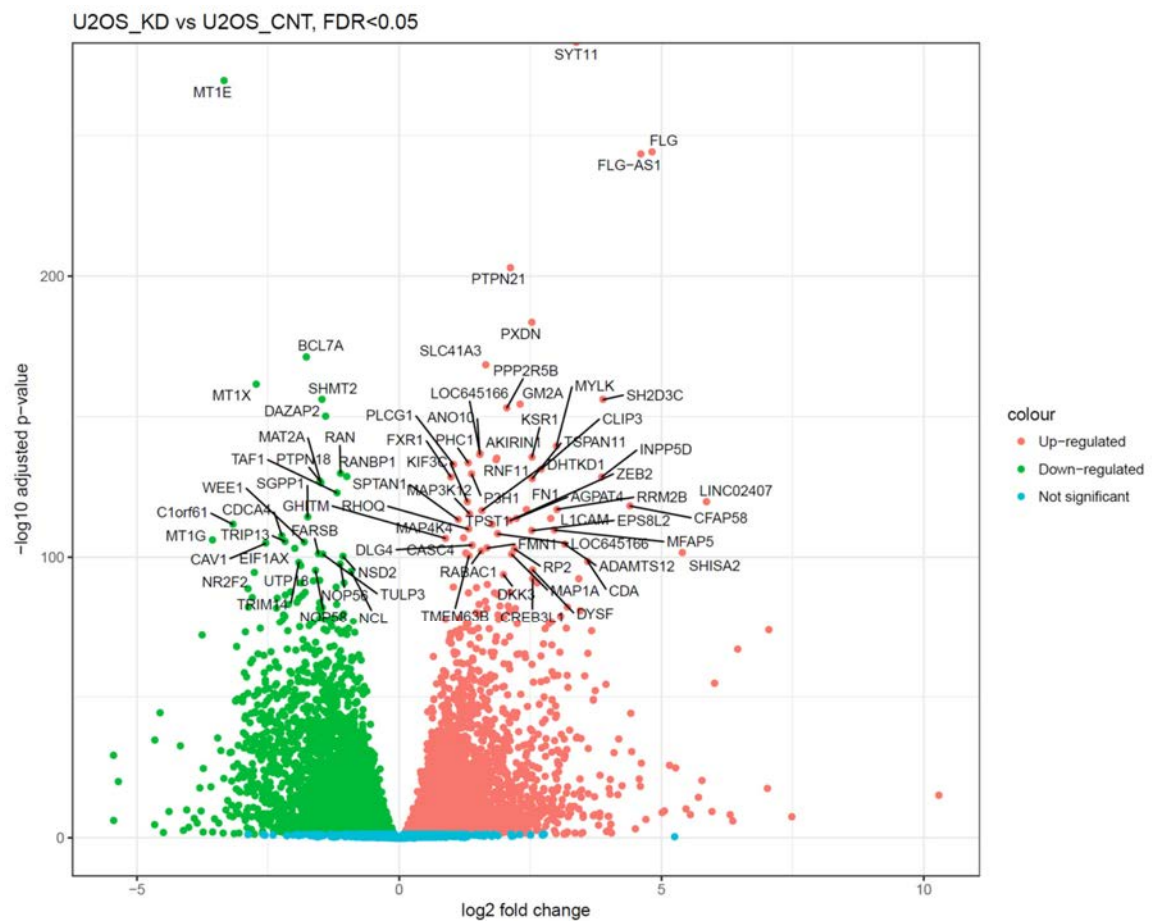
Figure 4.22 Volcanoplot of DE genes at depletion of NR2F2 in U2OS cell line.Top 80 DE genes are labelled.

To characterise the biological process that have changed as a result of NR2F2 depletion in U2OS cells, GO enrichment analysis was preformed using the DE genes (Total number= 852 DE genes, fold change >4, FDR <0.05) . This identified change in expression of several biological processes (Figure 4.23, and Appendix 4.7). The results of this analysis showed changes in biological processes that mainly encompasses genes related to mitotic nuclear division (downregulated: *BIRC5, SKA1/3, KIF15, CCNB1/2, ERCC6L, NUF, NDC80, CENPE, AURKA/B, BUB1, TTK, BUB1B PLK1, ESPL1, DSCC1, EREG, CDT1, CDK1, FEN1*; upregulated: *MISP, PDFFRB*), chromosome segregation (downregulated: *CENPM, ASPM, ESPL1, BUB1, PLK1, TTK, BUB1B, DSCC1, MK167, NEK2, CDCA5/8/20, PSRC1, ERCC6L, KIF14, KIF22*), spindle checkpoint (downregulated: *AURKB, NDC80, KNL, GSG2, SPDL1, TTK, BUB1B, BUB1, PLK, ESPL1, CDT1*), DNA replication (downregulated: *EXO1, RAD51, RFC4, PIF1, ORC1, MCM2-7 Complex, RRM1, E2F8, POLQ, CDC45, RFC3, PCLAF, GINS2, CHAF1A/B, BMP6, DTL, TICRR, GIS3, BRCA1, BRIP1, EME1, ESCO2, FEN1, CDK1, CDT1, CDC25A/C, DREG, DSCC1*; upregulated: *RRM2B, ACHE*), meiotic cell cycle (downregulated: *EXO1, RAD51, H2AFX, XRCC2, FANCA, MND1, EME1, BRIP1, PLK1, BUB1B, TTK, ESPL1, MKI67, AURKA, SGO1, KIF18A, CDC20, PTTG1, ASPM, TOP2A, DMC1, SGO2, FANCD2*; upregulated: *MOV10L1, REC8*). To reduce the number of genes that needed to be analysed, the clusterprofiler was used to extract the over-represented genes in GO enrichment of biological processes. The core of enriched biological processes is shown in (Figure 4.24).

GO enrichment analysis of molecular functions associated with downregulation of NR2F2 in U2OS, showed that mainly two molecular functions were changed (Figure 4.25); DNA-dependent ATPase activity (downregulated: *MCM4, BRIP1, PIF1, MCM2-7, CDC45, TOP2A, DSCC1, RFC3-4, POLQ, DMC1, RAD51, XRCC2, BLM, GINS2, CDC45*), and microtubule binding and motor activity (downregulated: *BRIC5, PSRC1, SYBU, NUSAP1, SKA1*, Kinesin Family proteins "*KIFC1, KIF2C, KIF4A , KIF11, KIF14, KIF15, KIF18A/B, KIF20A/B, KIF22*", *CDNP2, PLK1, RACGGAP1*; upregulated: *TRIM54, CLIP3, MAP1A, JAKMIP2*)
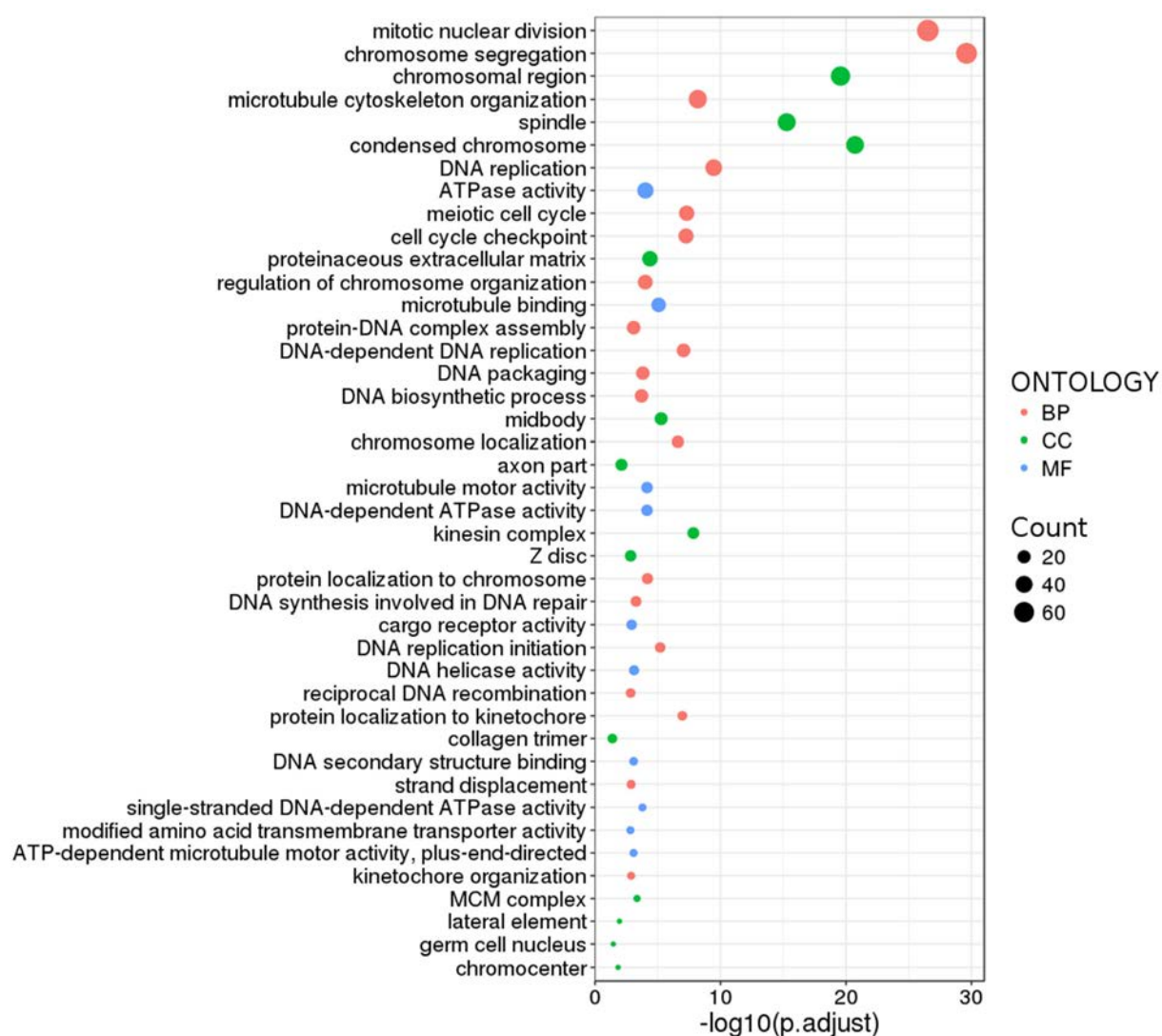
Figure 4.23 GO enrichment of DE genes upon NR2F2 downregulation in U2OS. Differently expressed genes in U2OS cell lines (fold change >4, FDR <0.05) was mapped to GO terms using David platform, where ALL GO terms were tested. The enrichment results were summarised by allowed medium similarity=0.7 using REViGO online tool (Supek *et al.*, 2011). GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.
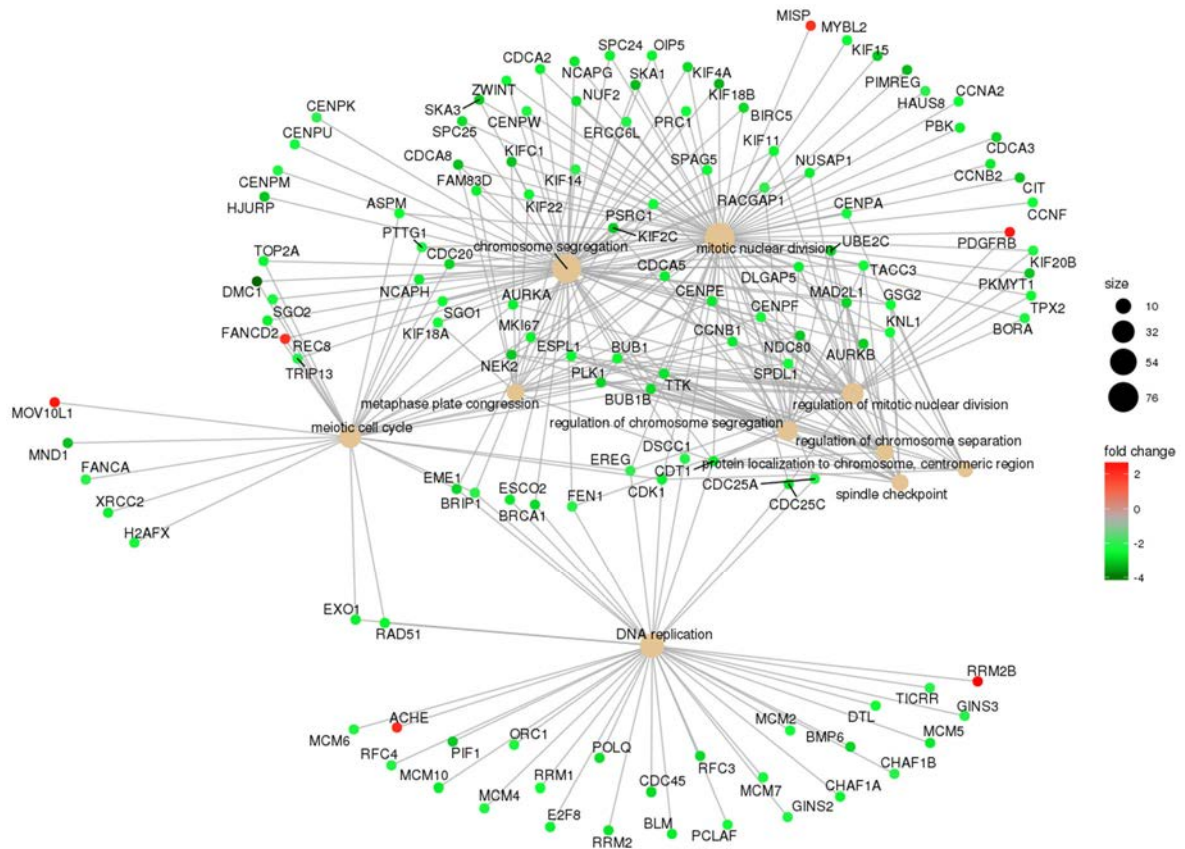
Figure 4.24 Visualisation of DE genes in most significantly enriched biological processes that associated with NR2F2 depletion in U2OS. It illustrates the biological complexities in which a gene may belong to multiple annotations, it also shows the terms linked to the most significant biological processes associated with genes that show to DE upon downregulation of NR2F2 in U2OS cells. Fold change >4, FDR <0.05.
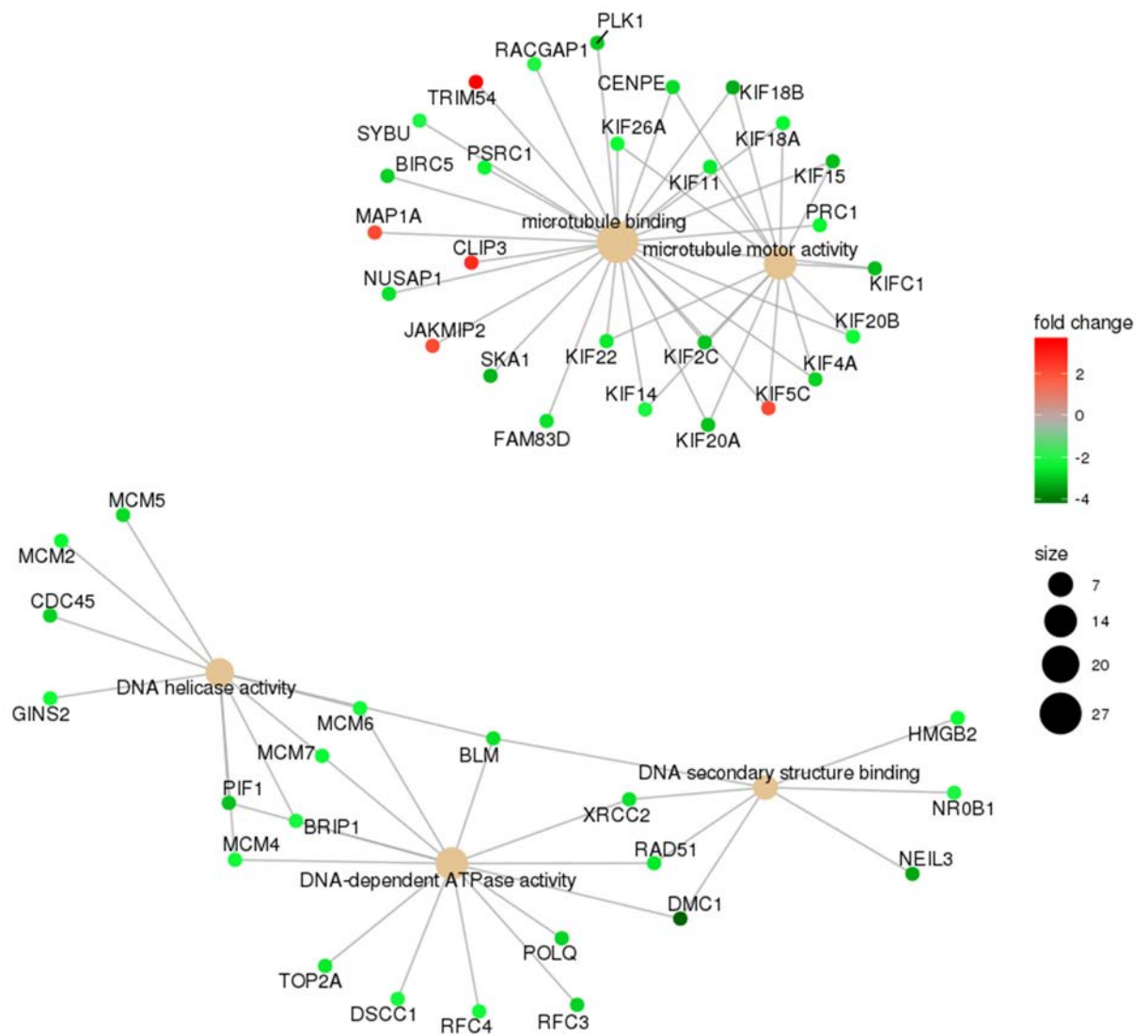
Figure 4.25 Visualisation of DE genes in most significantly enriched molecular function terms that associated with NR2F2 depletion in U2OS shows the most significant molecular function themes that associated to downregulation of NR2F2 in U2OS cells. Fold change >4, FDR <0.05.

In order to investigate the relationship between the ALT related genes, NR2F2-target genes and DE genes (fold change >2) in U2OS, triple intersection was carried out (Figure 4.26). The results showed 20 genes were overlapped between ALT related genes and DE genes (Table 4.7), and 65 genes are in common with NR2F2-targets (Table 4.8), while only *PSMC3IP/HOP2* was in common among the three lists. The 20 ALT related genes included RAD51-dependent homologous recombination genes (*RAD51, RAD51D*), meiotic recombination factors (*PSMC3IP/HOP2*, *MND1*), a group of Fanconi Anaemia genes (*FANCA, FANCD2, FANCM, BRIP1/FANCJ*), two topoisomerases (*TOP2A, TOP3A*), an RecQ helicase (*BLM*), an endonuclease that play role in DNA replication and synthesising the lagging strand gene (*FEN1*), an MRN complex subunit (*NBN*), and ssDNA binding protein (*RPA2*). In addition, the 65 NR2F2 target genes that showed DE in U2OS included some DNA replication genes such as *MCM5* (subunit of DNA helicase of MCM complex) and *RFC5* (subunit of replication factor C). Thus, it seems that NR2F2 is involved in regulating DNA replication in U2OS cells.
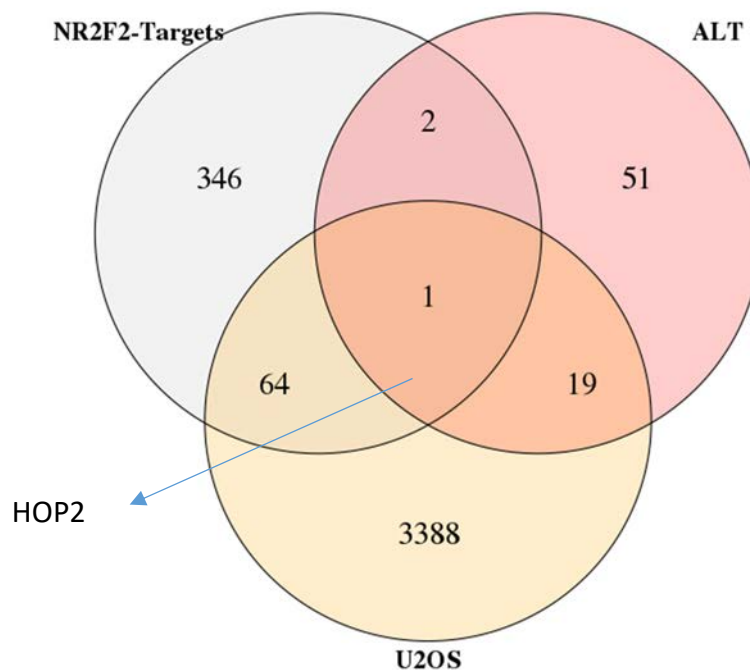
Figure 4.26 Intersection of DE genes in U2OS with ALT- related genes and NR2F2-trargets. Log2 fold change >1, FDR <0.05

Table 4.7 Intersection of ALT-related genes with DE genes in U2OS upon depletion of NR2F2. Log2 fold change >1, FDR <0.05.

| Ensambl Gene ID	baseMean	log2FoldChange	p value	p adj	Gene symbol	Full Name
ENSG00000144554	751.2677	-2.60905	5.87E-33	2.65E-31	FANCD2	Fanconi anemia complementation group D2
ENSG00000187741	2187.855	-2.10202	1.14E-29	3.86E-28	FANCA	Fanconi anemia complementation group A
ENSG00000187790	569.44	-1.31	6.55E-15	5.97E-14	FANCM	Fanconi anemia complementation group M
ENSG00000117748	1041.175	-1.42343	1.47E-28	4.43E-27	RPA2	replication protein A2
ENSG00000177302	1636.162	-1.13749	1.91E-28	5.73E-27	TOP3A	topoisomerase (DNA) III alpha
ENSG00000197299	612.342	-2.59808	1.39E-27	3.91E-26	BLM	Bloom syndrome RecQ like helicase
ENSG00000012048	1172.982	-2.77789	5.47E-26	1.35E-24	BRCA1	BRCA1, DNA repair associated
ENSG00000121211	266.3608	-3.10477	1.93E-24	4.11E-23	MND1	meiotic nuclear divisions 1
ENSG00000188486	5842.332	-2.2007	6.47E-22	1.13E-20	H2AFX	H2A histone family member X
ENSG00000168496	3285.559	-2.07731	9.00E-21	1.40E-19	FEN1	flap structure-specific endonuclease 1
ENSG00000137337	2957.557	-1.51273	9.51E-21	1.47E-19	MDC1	mediator of DNA damage checkpoint 1
ENSG00000139618	624.9101	-1.61237	8.50E-20	1.21E-18	BRCA2	BRCA2, DNA repair associated
ENSG00000104320	1904.869	-1.05625	1.05E-19	1.49E-18	NBN	nibrin
ENSG00000185379	619.0631	-1.05236	2.08E-18	2.62E-17	RAD51D	RAD51 paralog D
ENSG00000105011	693.707	-2.85047	1.61E-17	1.89E-16	ASF1B	anti-silencing function 1B histone chaperone
ENSG00000051180	368.9681	-2.44234	1.16E-15	1.15E-14	RAD51	RAD51 recombinase
ENSG00000129484	1924.718	-1.03058	2.42E-15	2.31E-14	PARP2	poly(ADP-ribose) polymerase 2
ENSG00000131470	268.5508	-1.87234	2.10E-13	1.66E-12	PSMC3IP (HOP2)	PSMC3 interacting protein
ENSG00000136492	998.4001	-1.99169	2.52E-12	1.76E-11	BRIP1	BRCA1 interacting protein C-terminal helicase 1
ENSG00000131747	9537.718	-2.48067	9.43E-09	4.44E-08	TOP2A	topoisomerase (DNA) II alpha

Table 4.8 Intersection of NR2F2-targets with DE genes in U2OS upon depletion of NR2F2. Log2 fold change >1, FDR <0.05.

Ensambl Gene ID	baseMean	log2FoldChange	p value	p adj	Gene symbol	Full Name
ENSG00000125462	610.2051	-3.19743	5.59E-119	1.01E-115	C1orf61	chromosome 1 open reading frame 61
ENSG00000102935	283.5264	2.884621	3.97E-65	1.25E-62	ZNF423	zinc finger protein 423
ENSG00000099365	316.9335	2.07789	2.30E-47	2.75E-45	STX1B	syntaxin 1B
ENSG00000132932	511.7102	2.703468	5.92E-36	3.32E-34	ATP8A2	ATPase phospholipid transporting 8A2
ENSG00000166123	4467.153	-1.98057	2.70E-30	9.52E-29	GPT2	glutamic--pyruvic transaminase 2
ENSG00000132879	1190.566	1.158195	1.20E-29	4.03E-28	FBXO44	F-box protein 44
ENSG00000144366	4433.001	1.071165	1.93E-28	5.76E-27	GULP1	GULP, engulfment adaptor PTB domain containing 1
ENSG00000139182	5152.642	1.061826	5.08E-27	1.37E-25	CLSTN3	calsyntenin 3
ENSG00000167522	14556.69	-1.21232	6.30E-27	1.69E-25	ANKRD11	ankyrin repeat domain 11
ENSG00000069956	2835.367	1.293594	1.45E-25	3.45E-24	MAPK6	mitogen-activated protein kinase 6

ENSG00000089847	146.1879	2.384474	1.90E-22	3.47E-21	ANKRD24	ankyrin repeat domain 24
ENSG00000100297	4674.249	-2.61975	4.19E-22	7.45E-21	MCM5	minichromosome maintenance complex component 5
ENSG00000235750	306.7452	-2.59438	2.61E-20	3.87E-19	KIAA0040	KIAA0040
ENSG00000075643	485.6794	-2.22593	3.69E-20	5.39E-19	MOCOS	molybdenum cofactor sulfurase
ENSG00000116661	884.7082	2.31269	9.94E-20	1.41E-18	FBXO2	F-box protein 2
ENSG00000140090	134.6197	-1.87344	1.83E-19	2.53E-18	SLC24A4	solute carrier family 24 member 4
ENSG00000197457	1381.322	-1.18607	1.89E-17	2.20E-16	STMN3	stathmin 3
ENSG00000163430	6222.504	1.421712	4.20E-17	4.75E-16	FSTL1	follistatin like 1
ENSG00000108852	894.8517	1.196367	5.53E-17	6.18E-16	MPP2	membrane palmitoylated protein 2
ENSG00000121957	1486.635	-1.73651	5.70E-17	6.37E-16	GPSM2	G-protein signaling modulator 2
ENSG00000188677	521.9604	-1.49451	2.95E-16	3.09E-15	PARVB	parvin beta
ENSG00000136859	326.2597	1.696921	2.21E-15	2.12E-14	ANGPTL2	angiopoietin like 2
ENSG00000177465	215.3652	-1.70024	3.09E-15	2.92E-14	ACOT4	acyl-CoA thioesterase 4
ENSG00000203877	56.65877	-2.93152	1.17E-14	1.04E-13	RIPPLY2	ripply transcriptional repressor 2
ENSG00000070759	347.9056	1.225537	3.75E-14	3.20E-13	TESK2	testis-specific kinase 2
ENSG00000173638	1060.452	-1.36803	9.94E-14	8.12E-13	SLC19A1	solute carrier family 19 member 1
ENSG00000008130	2352.244	-1.00274	1.05E-13	8.55E-13	NADK	NAD kinase
ENSG00000158555	1691.13	1.44491	1.59E-13	1.27E-12	GDPD5	glycerophosphodiester phosphodiesterase domain containing 5
ENSG00000131470	268.5508	-1.87234	2.10E-13	1.66E-12	PSMC3IP	PSMC3 interacting protein
ENSG00000213347	1089.92	-1.71402	3.30E-13	2.56E-12	MXD3	MAX dimerization protein 3
ENSG00000116544	392.0044	1.089932	3.56E-13	2.75E-12	DLGAP3	DLG associated protein 3
ENSG00000176890	2253.868	-1.33038	3.74E-13	2.87E-12	TYMS	thymidylate synthetase
ENSG00000118898	74.66078	2.065579	8.36E-13	6.18E-12	PPL	periplakin
ENSG00000198205	1641.818	1.604545	4.78E-12	3.25E-11	ZXDA	zinc finger, X-linked, duplicated A
ENSG00000132563	382.8398	1.143674	1.06E-11	6.94E-11	REEP2	receptor accessory protein 2
ENSG00000170089	440.2057	-1.11389	1.73E-11	1.11E-10	LOC728554	THO complex 3 pseudogene
ENSG00000185761	558.3325	1.389001	3.33E-11	2.06E-10	ADAMTSL5	ADAMTS like 5
ENSG00000089558	61.46915	-2.18379	1.05E-10	6.15E-10	KCNH4	potassium voltage-gated channel subfamily H member 4
ENSG00000174529	72.69154	1.710668	1.17E-10	6.83E-10	TMEM81	transmembrane protein 81
ENSG00000178726	116.837	-2.08056	1.42E-10	8.24E-10	THBD	thrombomodulin
ENSG00000140044	1366.307	-1.21742	1.50E-10	8.73E-10	JDP2	Jun dimerization protein 2
ENSG00000149929	337.029	-1.44084	1.66E-10	9.62E-10	HIRIP3	HIRA interacting protein 3
ENSG00000204128	208.978	-1.31203	1.79E-10	1.03E-09	C2orf72	chromosome 2 open reading frame 72
ENSG00000111445	608.7437	-1.32039	2.10E-09	1.06E-08	RFC5	replication factor C subunit 5
ENSG00000142765	290.8645	-1.04117	3.73E-09	1.84E-08	SYTL1	synaptotagmin like 1
ENSG00000138621	185.0986	-1.12376	3.92E-09	1.93E-08	PPCDC	phosphopantothenoylcysteine decarboxylase
ENSG00000162520	30.4848	2.589262	5.83E-09	2.81E-08	SYNC	syncoilin, intermediate filament protein
ENSG00000101230	167.9533	1.149013	1.69E-08	7.68E-08	ISM1	isthmin 1
ENSG00000170579	92.21989	-1.26797	2.15E-08	9.63E-08	DLGAP1	DLG associated protein 1
ENSG00000184678	1923.199	1.44556	1.54E-07	6.18E-07	HIST2H2BE	histone cluster 2 H2B family member e
ENSG00000103723	137.0842	-1.06781	2.07E-07	8.13E-07	AP3B2	adaptor related protein complex 3 beta 2 subunit
ENSG00000130347	122.2309	-1.09103	2.54E-07	9.86E-07	RTN4IP1	reticulon 4 interacting protein 1

ENSG00000118322	36.48111	2.136534	3.54E-07	1.34E-06	ATP10B	ATPase phospholipid transporting 10B (putative)
ENSG00000177807	47.10988	1.596174	2.48E-06	8.37E-06	KCNJ10	potassium voltage-gated channel subfamily J member 10
ENSG00000139531	125.564	1.152468	2.55E-06	8.57E-06	SUOX	sulfite oxidase
ENSG00000019169	805.6424	2.799769	3.39E-05	9.55E-05	MARCO	macrophage receptor with collagenous structure
ENSG00000116663	55.86403	1.066421	0.000175	0.000441	FBXO6	F-box protein 6
ENSG00000158055	41.74765	1.343401	0.000224	0.000554	GRHL3	grainyhead like transcription factor 3
ENSG00000140557	74.29115	-1.17714	0.000241	0.000594	ST8SIA2	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2
ENSG00000113319	38.70975	1.27358	0.00051	0.001186	RASGRF2	Ras protein specific guanine nucleotide releasing factor 2
ENSG00000054356	5470.262	1.947802	0.000779	0.001753	PTPRN	protein tyrosine phosphatase, receptor type N
ENSG00000166578	57.7997	1.195363	0.001157	0.002526	IQCD	IQ motif containing D
ENSG00000183018	804.8771	-1.98066	0.001292	0.002796	SPNS2	sphingolipid transporter 2
ENSG00000277182	41.54543	-1.03252	0.001534	0.003281	LOC100287808	uncharacterized LOC100287808
ENSG00000103184	32.11846	1.255009	0.002969	0.006034	SEC14L5	SEC14 like lipid binding 5

4.4 Discussion

In chapter 3, I showed that NR2F2 was expressed at a significantly higher level in the WV cell line (~ 2 fold increased) compared to the other ALT+ cell lines, suggesting an amplification might arise from gains of functional copies of this gene in WV cell line. The high level of NR2F2 expression in the WV cell line did not result in preferential localisation at telomeres. Analysis of XpYp telomere mutations in this cell line showed loss of variant or degenerate repeats compared to the progenitor allele and replacement with canonical $(TTAGGG)_n$ repeats (Mendez-Bermudez *et al.*, 2012).

In this chapter, NR2F2 was downregulated in WV cell line, and other ALT+ cells to identify the poteintal roles of NR2F2 in these cells. Depletion of NR2F2 (87%) in WV was associated with an increase in the percentage of dead cells, an increase in C-circle abundance, while no significant change in the frequency of APBs. Furthermore, a moderate change in G1 and reduction in proportion of cells at the S phase, while no change in G2 was observed in this cell line. This suggests that NR2F2 regulates the expression of growth factors and genes involved in cell survival in this cell line. A possible explanation for the elevated level of C-circles upon NR2F2 depletion in these cells is because of NR2F2's role in regulating the expression of some DNA helicases that are required for resolving the replication stress at telomeres causing an increased level of replication stress in these cells.

More extreme effects were observed after nearly complete downregulation of NR2F2 (94%) in U2OS. The cell cycle was arrested in G1 and G2, there was a significant increase in C-circles (> 1 fold change) and substantial increase in frequency of nuclei with APBs (~ 4 fold change). The function of APBs in ALT+ cells is unclear, but it was found that these structures are assembled and disassembled in a cell cycle dependent manner and are mainly visible in G2/M (Grobelny *et al.*, 2000). However, the use of methionine restriction to enlarge APBs in the IIICF/c (ALT-positive, Li-Fraumeni Syndrome) cell line showed that they are not restricted to G2/M, but also can be formed at cell cycle arrest in G0/G1 (Jiang *et al.*, 2007). This may suggest that the substantial increase in the frequency of nuclei with APBs is because most U2OS cells were arrested in G0/G1 or G2/M phases of cell cycle.

Different effects were observed upon NR2F2 downregulation (84%) in the WI38VA13_2RA cell line. There was a significant reduction in the frequency of APBs that correlated with the decrease in the percentage of cells in G2/M. Nevertheless, no change in the level of C-circles was observed in these cells. These observations are comparable to a previous study in which no significant change in C-circle was shown after NR2F2 depletion (Conomos *et al.*, 2012).

In order to understand the regulatory roles of NR2F2 in ALT mechanism, differential expression analysis by RNA-seq was performed on cells with depleted NR2F2. Overall, the DE analysis showed a more significant effect of NR2F2 depletion in U2OS than in WV or WI38VA13_2RA, as shown by the percentage of DE genes at log2 fold change>1 (Table 4.4). Some similar effects were found in WV and WI38VA13_2RA, these centred on changes in the expression of genes associated with GO terms related to blood vessel morphology, protein kinases, cell growth, and cell development and differentiation, and to a lesser extent on cell cycle genes. No change in the expression of ALT-related genes or any genes that are known to be involved in DNA repair pathways was observed in WV or WI38VA13_2RA, which suggests no transcriptional targets of ALT related genes for NR2F2 in these two cell lines. Conomos et al 2014 showed that downregulation of NR2F2 in WI38VA13 results in a reduction in ALT markers such as the number of APBs per cells has changed from 22 to 11, and a reduction in recruitment of NuRD-ZNF827 at telomeres. However, there were no direct changes in the expression of genes that encode NuRD complex and ZNF827.

Furthermore, the fact that both WV and WI38VA13_2RA are SV40-immortalised fibroblasts with a similar expression profile could explain the limited requirement of NR2F2 in regulating the cell cycle (Figure 4.27 and Appendix 4.8). The SV40 large T antigen in these cells not only inactivates p53 and Rb tumour suppressor proteins, but also regulates the cell cycle by activating E2F transcriptional factors (Saenz-Robles *et al.*, 2007); DNA replication via regulating the expression of MCM helicase complex and RFC5 (Deeb *et al.*, 2007). Furthermore, the elevated level of NR2F2 in the WV cell line seems to be required for sustaining cell proliferation signalling and for resisting cell death, while no role in ALT mechanism was observed.

By contrast, depletion of NR2F2 in U2OS resulted in downregulation of 20 of the 73 ALT-related genes. The effect of NR2F2 depletion in U2OS included many other genes that were associated with GO terms related to DNA replication (such as *MCM* complex, *RFC3/4, POLQ*), chromosome segregation, spindle checkpoint and meiotic cell division. This suggests that NR2F2 plays an important role in regulating the expression of genes that involved DNA replication and DNA repair mechanisms, where some of them seems to be related to the ALT mechanism. The ALT related genes included the RAD51-dependent homologous recombination genes (*RAD51, RAD51D, XRCC2*, and *PSMC3IP,MND1*), a group of Fanconi Anaemia genes (*FANCA, FANCD2, BRIP1/FANCJ* and *FANCM*), topoisomerases (*TOP2A, TOP3A*), RecQ helicases (*BLM*), and an endonuclease (*FEN1*).

However, the reduction in the expression of RPA2 (subunit of the heterotrimeric RPA complex) upon NR2F2 depletion in U2OS seems might be the main reason for the observed cell cycle arrest in G2 and probably the increased level of C-circles and elevated frequency of cells with APBs. It was shown by Grudic et al. (2007) that loss of the single-stranded DNA-binding RPA protein in human ALT+ cells, but not in TEL+ cells, causes increased exposure of single-stranded G-rich telomeric DNA, cell cycle arrest in G2, accumulation of single-stranded telomeric DNA within APBs.

In summary, this work has shown three different patterns of effects on ALT markers and the cell cycle upon NR2F2 depletion. First, no significant changes in ALT markers nor in the percentage of cells in G2/M was shown in WV. Second, a significant reduction in the frequency of APBs that correlated with a decrease in percentage of cells in G2/M was shown in WI38VA13, but no change in C-circles. Third, contradictory effects; significant increase in ALT markers combined with cell cycle arrest at G1 or G2/M in U2OS. NR2F2 knockdown resulted in significant changes in expression of genes involved in DNA replication and repair and with roles in the ALT mechanism in U2OS, but not in WV or WI38VA13_2RA. These changes in gene expression were consistent with the observed changes in ALT phenotypic markers and cell cycle disruption. The depletion of NR2F2 in U2OS is also likely to have a major impact on DNA replication.
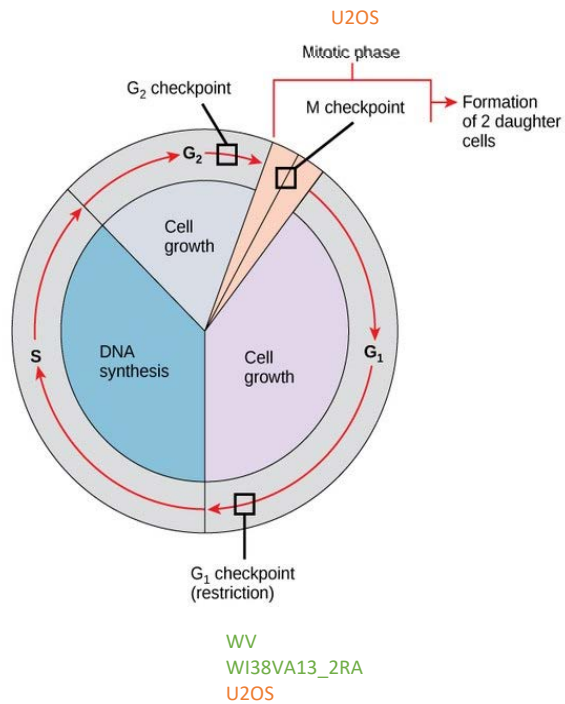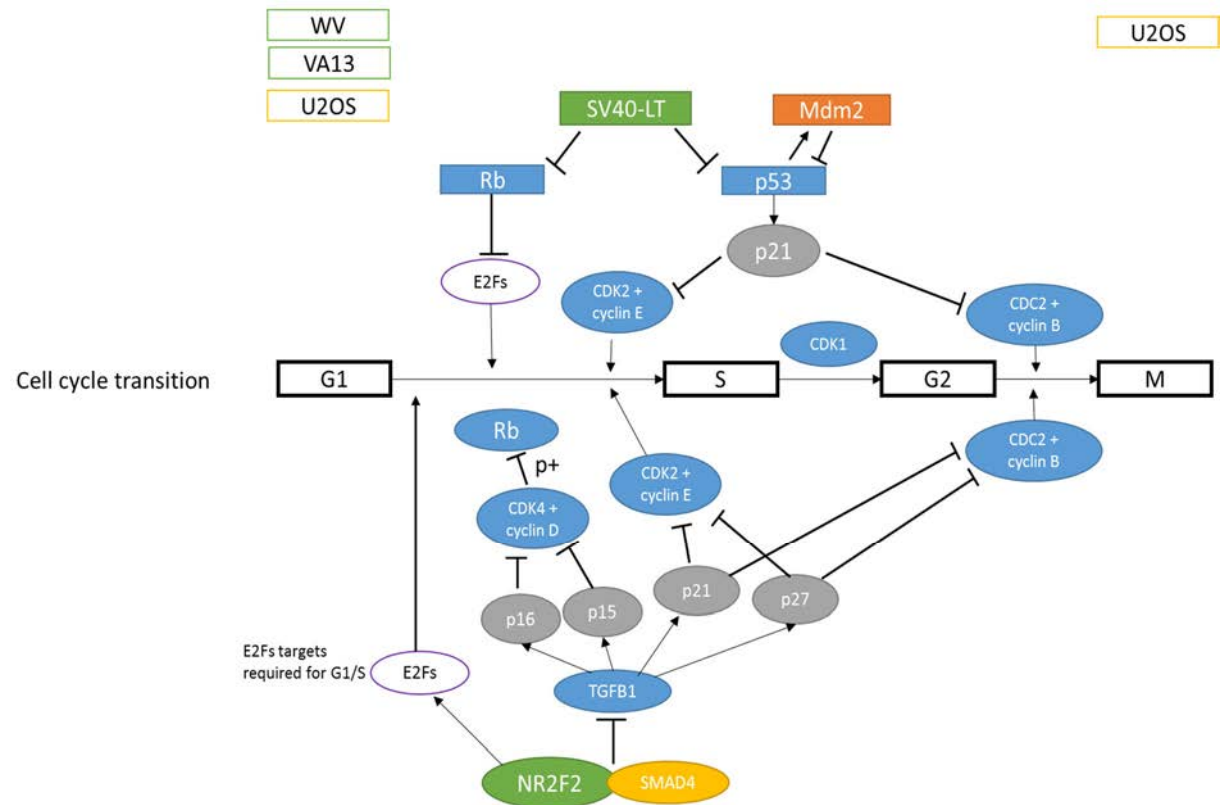
Figure 4.27 Model of NR2F2 as cell cycle regulation in ALT+ cells. (A) the checkpoints where the changes in cell cycle were observed are shown (B) NR2F2 and SV40-LT target genes sharing the same targets that regulate the G1 and G2 checkpoints.

Chapter 5: *ATRX* knockout in pre-crisis SV40-transformed fibroblast using CRISPR-Cas9 gene-editing

5.1 Introduction

5.1.1 Loss of *ATRX* in cells and tumours that use ALT

ATRX (Alpha-Thalassemia/mental Retardation X-linked) is an ATP-dependent helicase belongs to the SWI/SNF family of chromatin remodelling proteins. *ATRX* germline mutations are associated with a complex genetic disorder termed alpha-thalassemia/mental retardation X-linked (ATR-X) syndrome in which ATRX, as a part of the chromatin remodelling complex, participates in regulation of transcription of a set of genes, where *HBA1* (Haemoglobin, alpha 1) is the best characterised (Gibbons *et al.*, 1992). In contrast, *ATRX* somatic mutations and reduced expression are found in variety of cancers, especially those that use the ALT mechanism (Heaphy *et al.*, 2011a; Schwartzentruber *et al.*, 2012).

The full-length functional ATRX protein is 2492aa (~ 280 kDa), comprising two highly conserved domains: the ATRX-DNMT3-DNMT3L (ADD) domain and the SWI/SNF helicase domain, as well as the C-terminal helicase (Park *et al.*, 2004). The ADD domain contains a plant homeodomain (PHD) finger that recognises and binds to two of histone modifications, H3K9me3 and H3K4me0 (Iwase *et al.*, 2011; Eustermann *et al.*, 2011). ATRX is also known to show physical interactions with several proteins (Figure 5.1). ATRX interacts with DAXX (Death-domain associated protein) (Tang *et al.*, 2004), HP1α (Chromobox Homolog 5/ HP1 Alpha Homolog) (Lechner *et al.*, 2005), and EZH2 (Enhancer of Zeste Homolog 2) (Cardoso *et al.*, 1998), MECP2 (Methyl-CpG Binding Protein 2) (Nan *et al.*, 2007), and the histone variant (macroH2A1) (Ratnakumar *et al.*, 2012).
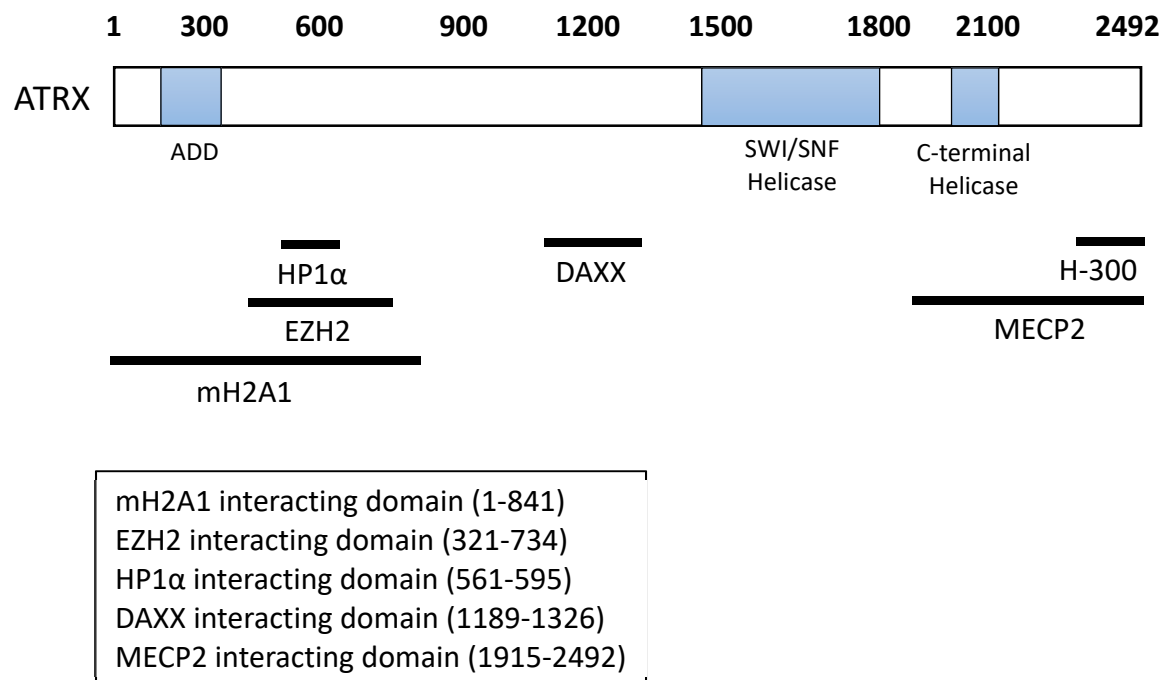
Figure 5.1 Structure of the human ATRX protein. The full length of ATRX protein is 2492aa (~280 kDa) featuring the ATRX-DNMT3-DNMT3L (ADD) domain, the SWI/SNF helicase, and the C-terminal helicase. The regions through which ATRX interacts with macroH2A1, EZH2, HP1α, DAXX, and MECP2 are shown. The C-terminal location of the immunogenic sequence, H-300, used to generate the polyclonal ATRX-H-300 antibody is shown.

A well-characterised function of ATRX is to work with its partner DAXX, as a histone chaperone complex for the histone variant H3.3, in order to maintain the heterochromatin status at repetitive regions (Figure 5.2). These repetitive regions include telomeres and pericentromeric repeats (McDowell *et al.*, 1999; Drane *et al.*, 2010; Goldberg *et al.*, 2010; Wong *et al.*, 2010; Lewis *et al.*, 2010), some retrotransposon families (Elsasser *et al.*, 2015; Sadic *et al.*, 2015), and ribosomal DNA (rDNA) repeat arrays (McDowell *et al.*, 1999; Udugama *et al.*, 2018). In a different mechanism, loss of ATRX contributes to an increased incorporation of macroH2A1 at the HBA1 gene cluster on the subtelomeric region of chromosome 16, and decreased expression level of α-globin. This observation suggested that ATRX is a negative regulator of macroH2A1, and partially explained the pathological changes in haemoglobin levels that are associated with germline mutations in ATR-X syndrome, particularly mutations in exonic regions

that coding for N-terminal, and ADD domain (Ratnakumar *et al.*, 2012). In contrast, the germline mutations that clustered at the C-terminal helicase domain showed disruption of the ATRX- MECP2 complex leading to pathological changes that are associated with mental retardation (Nan et al., 2007). It was found that ATRX interacts with EZH2 (Cardoso *et al.*, 1998). EZH2 is a member of the Polycomb group (PcG) of proteins, and the catalytic subunit of the PRC2-EZH2 complex. This complex participates in histone methylation of H3K27 that results in transcriptional repression of target genes (Cao *et al.*, 2002), which are also known to drive X- chromosome inactivation during female development in mammals. Moreover, it was found that EZH2 controls the activity of DNA methyltransferases (DNMTs) at the promotor regions of EZH2-repressed genes (Vire *et al.*, 2006). The global effect of ATRX is to mediate activity of the PRC2-EZH2 complex in spreading the H3K27 methylation and in chromosome X-inactivation in mouse embryonic fibroblasts (Sarma *et al.*, 2014).

Somatic mutations of ATRX (or its partner DAXX) are significantly associated with suppression of the ALT mechanism in cell lines and tumours. For example, ATRX was mutated or absent in 86% (19/22) of ALT+ cell lines (Lovejoy *et al.*, 2012), mutated in ~ 80% (37/47) of ALT+ paediatric glioblastoma (Schwartzentruber *et al.*, 2012), and 100% (19/19) of ALT+ pancreatic neuroendocrine tumours (Heaphy *et al.*, 2011a). The role of ATRX in suppression of ALT is unclear (Clynes *et al.*, 2015) but re-expression of ATRX in ALT+ cells resulted in a reduction in replication fork stalling (a possible trigger for telomere maintenance by HR) by resolving G-quadruplexes and reduction in all ALT markers in DAXX/H3.3 dependent manner (Clynes *et al.*, 2015). Another study showed that absence of functional ATRX suppresses the resolution cohesion between sister telomeres that normally occurs prior to M phase, resulting in persistent telomere cohesion that was suggested to facilitate ALT through promoting intra-chromosomal telomeric recombination between sister telomeres in ALT+ cells (Ramamoorthy and Smith, 2015). Recently, it was found that loss of ATRX facilitates elongation of lagging strand overhangs in SW39 ALT+ cells (Min *et al.*, 2017). Nevertheless, the depletion of ATRX in human normal primary cells was not sufficient for immortalisation nor for activation of ALT in the presence of active M1 checkpoint (Lovejoy *et al.*, 2012; Napier *et al.*, 2015).
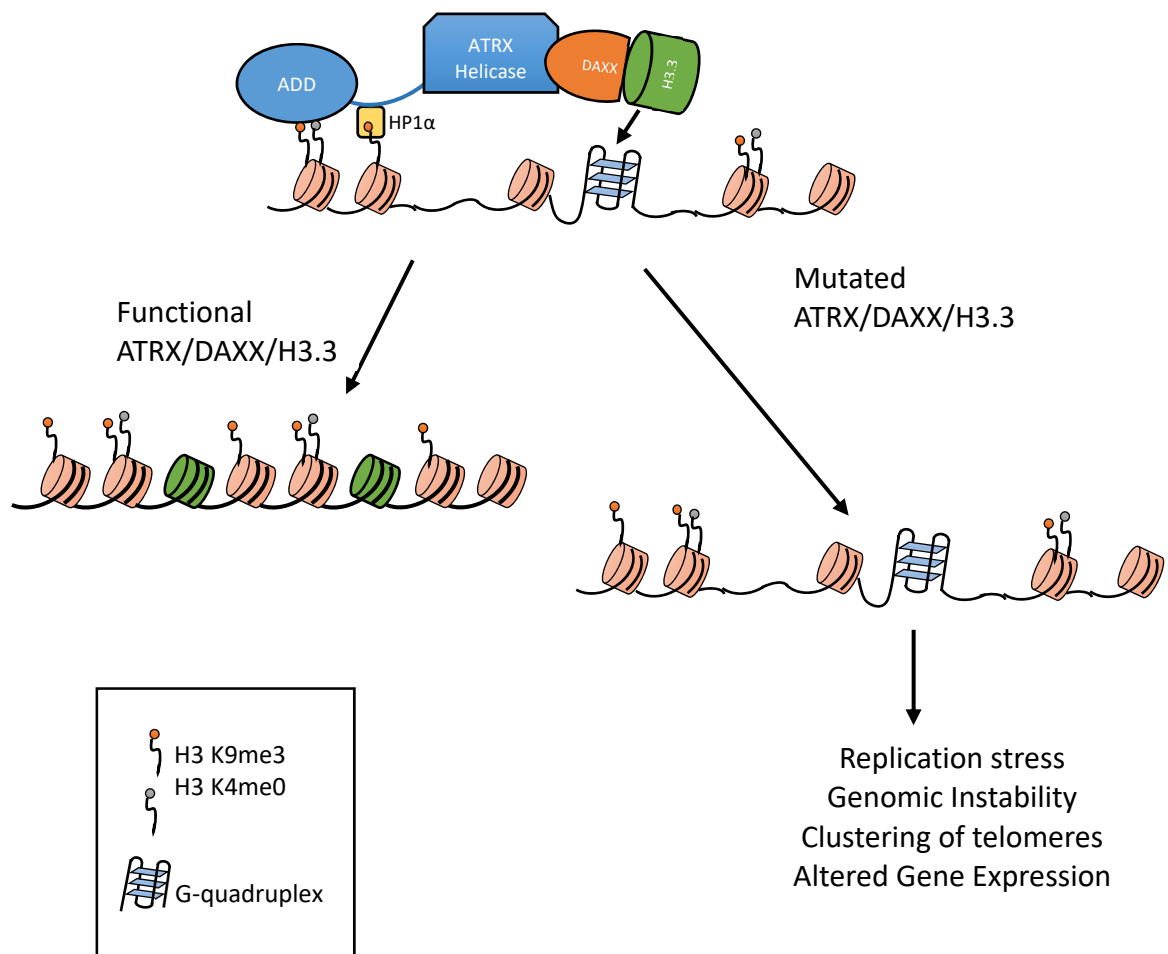
Figure 5.2 Model for the role of ATRX in maintaining genomic stability by regulation of heterochromatin in repetitive regions. ATRX works together with DAXX to deposit the histone variant H3.3 at repetitive regions, such as telomeres and pericentromeric regions. ATRX is also implicated in resolving the DNA secondary G-quadruplexes. Adopted from Eustermann *et al.*, 2011.

5.2 Aims

Acquisition of somatic mutations in ATRX or its partner DAXX is strongly associated with activation of the ALT mechanism in cell lines (Lovejoy *et al.*, 2012), and in ALT+ tumours (Schwartzentruber *et al.*, 2012; Heaphy *et al.*, 2011b). Consequently ATRX and DAXX appear to be suppressors of ALT and therefore loss of ATRX expression facilitates but is not regarded as being sufficient for the immortalisation of fibroblasts that have inactivated the senescence checkpoint (M1).

To investigate the relationship between ATRX inactivation and cell immortalisation and to identify other changes that may be required, CRISPR-Cas9 genome editing was used to mutate the ATRX gene. The inactivating ATRX mutations were generated in a pre-crisis SV40- transformed clonal fibroblast cell line that was otherwise normal. ATRX disruption was confirmed using western blot, IF-staining, and sequencing. Subsequently the telomere maintenance mechanism (TMM) was characterised in six independent clonal populations that survived crisis using variety of methods: (1) Assaying for ALT activation using ALT-associated phenotypes such APBs analysis, and telomere length analysis; (2) investigating TERT expression using RNA-seq. Cells were collected at different time points following emergence from crisis for further analyses described in Chapter 6.

5.3 Results

5.3.1 Designing a CRSIPR-Cas9 vector to target the ATRX gene in a pre-crisis SV40-transformed cell line.

5.3.1.1 Confirming ATRX expression in the pre-crisis SV40-transformed cells

The CCD18LU (primary fibroblast cell line derived from lung) had been transformed previously in Dr Royle's laboratory by transfection with a pSV3neo plasmid, G418 antibiotic selection and isolation of single colonies with stable SV40-LT expression. One of these pre-crisis SV40-trasnformed clones, CCD18LU/TCL1 (abbreviated to TCL1) was selected for ATRX editing using CRISPR-Cas9 system. TCL1 showed morphological changes compared to the parental CCD18LU cells and rapid growth. Expression of ATRX was confirmed in TCL1 using western blot analysis (Figure 5.3 A) and immunofluorescence staining (Figure 5.3 B). Western blot detection of ATRX was carried out using the polyclonal antibody ATRX-H-300 (Santa Cruz), first used by Garrick et al. in (2004) to investigate the expression of truncated isoforms of ATRX that lack the functional SWI/SNF helicase domain. The immunogenic sequence detected by the ATRX-H-300 antibody comprises 300aa (2193-2492) near the C-terminus of the ATRX protein, which made it a suitable choice to investigate inactivating mutations following CRISPR-Cas9 cleavage and NHEJ mediated frameshift/nonsense mutations.

The western blot showed presence of full length ATRX in TCL1 (Figure 5.3), and that a minimum of 30μg of total protein is required to detect the ATRX protein. Although no band was detect above of 180 kDa in the U2OS negative control (Lovejoy *et al.*, 2012), but some of non-specific bands were detected below 245 kDa in HT1080, BR149, and TCL1 (Figure 5.3 A). The same antibody was use for IF staining for ATRX in TCL1 with the same cell line controls. IF staining showed that ATRX foci are present in interphase nuclei of HT1080, BR149, and in TCL1, but not in U2OS (Figure 5.3 B). These results confirmed that ATRX is expressed in the pre-crisis SV40-transfected-TCL1 cells and therefore CRISPR-Cas9 mediated ATRX knockout was pursued.
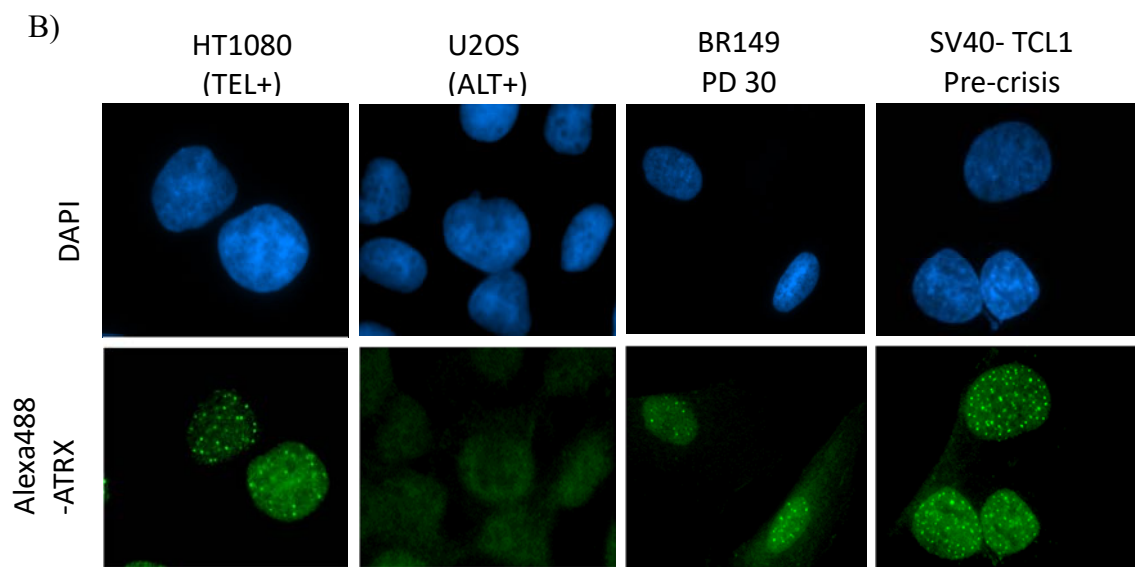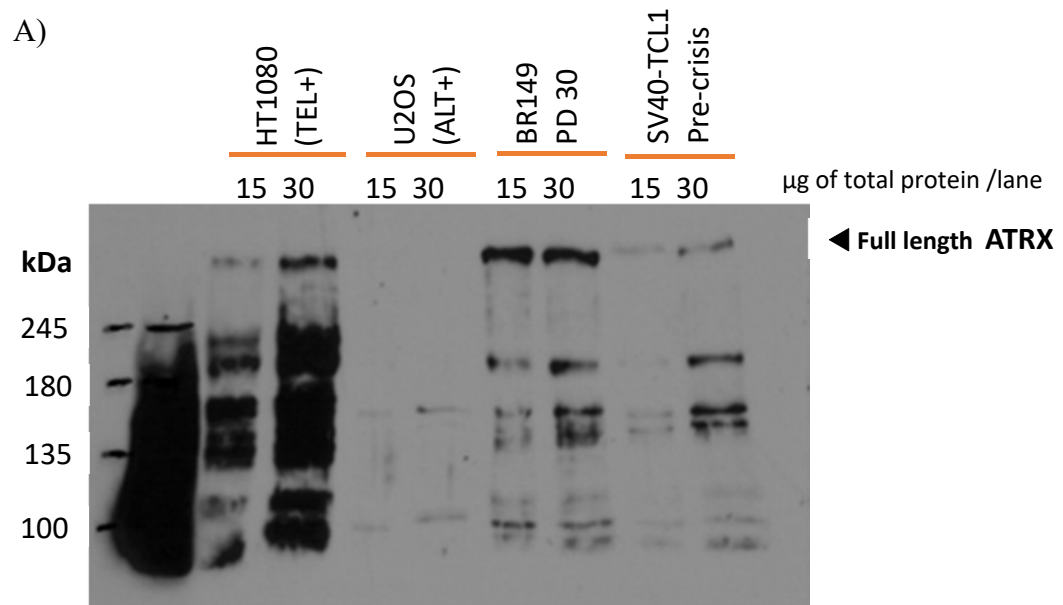
Figure 5.3 ATRX is expressed in pre-crisis SV40-T-antigen-tranfected-TCL1 cells. (A) Western blot detection of full length ATRX using the polyclonal ATRX-H-300 antibody from Santa Cruz. (B) Indirect immunofluorescence staining of ATRX in interphase nuclei using anti-IgG-Alexa-488 (green) to detect the localisation of the ATRX-H-300 primary antibody in interphase. The nuclei are counterstained with DAPI (blue). HT1080 (TEL+) and BR149 (mortal primary fibroblast) are positive controls, and U2OS (ALT+) was used as a negative control.

5.3.1.2 Selecting the CRISPR target sequences for editing of *ATRX*

The human *ATRX* gene spans ~ 281kb on the X chromosome (q13.3), and consists of 35 exons (Figure 5.4). The *ATRX* exon 9 of chosen as the target for CRISPR-Cas 9 mediated editing for several reasons (Figure 5.4 A). First, it is the position of exon 9 in *ATRX* gene, which is likely to increase the possibility of complete disruption by CRISPR-Cas9 mediated editing, and may also trigger NMD, which results in mRNA termination (Popp and Maquat, 2016). Second, exon 9 is the largest exon in the *ATRX* gene (3201 bp) and it codes for 1023 aa (222-1245), which partially includes the ADD domain, which is essential for ATRX function as a chromatin remodelling protein (see Appendix 5.5 for aa annotation of ATRX). Third, around 50% of *ATRX* inactivating mutations in gliomas are clustered in exon 9 (Figure 5.4 B) (Jiao *et al.*, 2012). Consequently, *ATRX* exon 9 was selected as the best target for efficient *ATRX* targeting and to simulate the mutations that are associated with ALT activation in tumours.

Overall, Cas9 can tolerate up to five nucleotide mismatches within a 20 nt guide RNA sequence and still cleave the target sequence (Fu *et al*. 2013; Hsu *et al*. 2013; Fu *et al*. 2014). To reduce the possibility of non-specific editing by CRISPR-Cas9 as a result of tolerated off-targets elsewhere in human genome, an intensive bioinformatics analysis was preformed to identify unique guide sequences that would target only ATRX exon 9. List of 112 potential ATRX exon 9 target sequences was assembled by combining the potential candidates from three online CRISPR target sequence design tools (CHOPCHOP http://chopchop.cbu.uib.no/ (Montague *et al.*, 2014); Blue Heron BIOTECH - ORIGENE's CRISPR target sequence design tool https://www.blueheronbio.com/ ; GeneArt CRISPR Search and Design), and Design, a ThermoFisher scientific tool "https://www.thermofisher.com/"). The 112 potential targets is the total number found by these tools in both plus and minus strands and after removing duplicates.

A)

ATRX
Chromosome X: 77,504,878-77,786,269
Coding exons: 35
Gene Size: 281. 39 kb

Exon 9

B)

ADD   SNF2 Helicase   C-terminal
                              Helicase
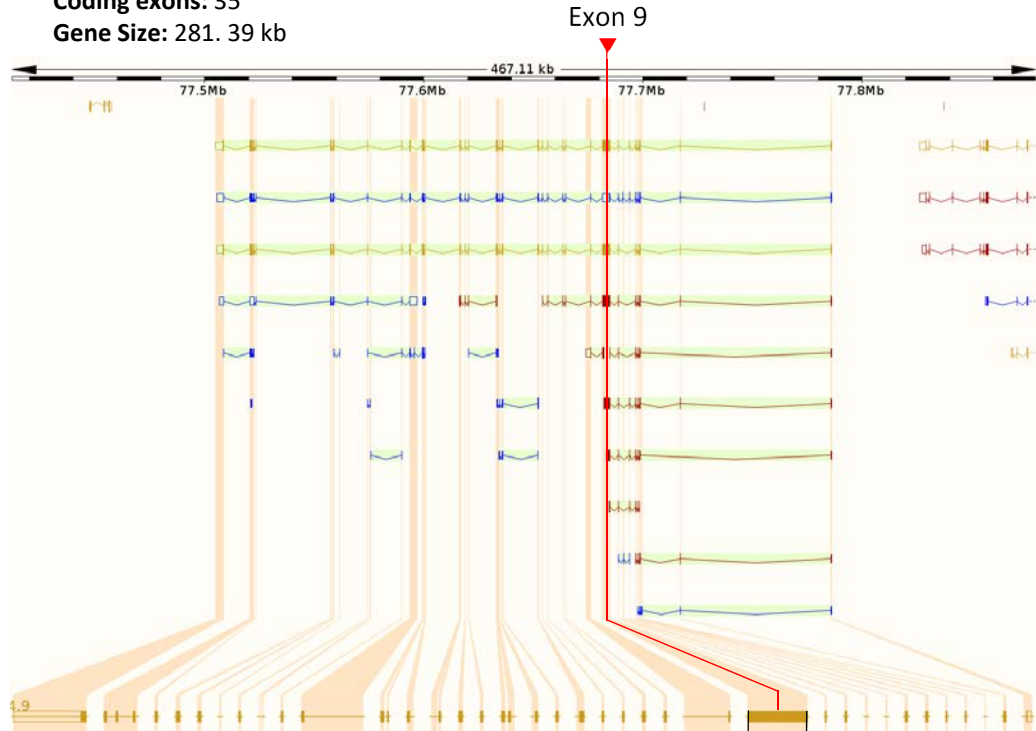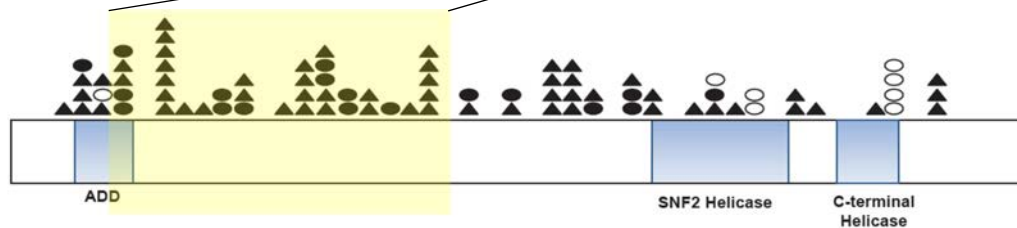
Figure 5.4 Location of *ATRX* exon 9 in splice variants and positions of ATRX mutations in ALT+ gliomas. (A) ATRX splice variants (downloaded from Ensmbl genome browser - Human (GRCh38). The red line marks Exon 9. (B) Distribution of ATRX mutations in ALT+ gliomas. Frameshift mutations are represented by shaded triangles, nonsense mutations correspond to shaded circles, while missense mutations were represent by empty circles. The yellow highlighted part represents the protein sequence encoded by Exon 9 (Jiao et al., 2012).

Short sequence blast against the human genome (using Blast+ v2.5.0 tool, https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST) was used to characterise the list of 20-nt length potential target sequences.

The blast results were filtered out by counting the number of hits for each potential target sequence in the human genome with zero to four mismatches. Among them, 92 target sequences were filtered out as a unique hit in case of zero mismatches (100% identity of 20-nt), 54 potential targets sequences were filtered out as the unique hit in case of one mismatch in variable positions in 20-nt, and only 10 target sequences were demonstrate an acceptable level of specify to be unique hit in the human genome in case of 3 mismatches (Figure 5.5). The two targets that are unique in exon 9 up to 3 mismatches were selected in this work. The two selected target sequences had 50% GC content, were not found in repetitive regions, and did not overlap with any known common SNPs. These CRISPR target sequences are CRISPR-TS-1 "TCTACGCAACCTTGGTCGAA" on the anti-sense strand, and CRISPR-TS-2 "GGGTGTAGTCTTTACACGTG" on sense strand (Figure 5.6).
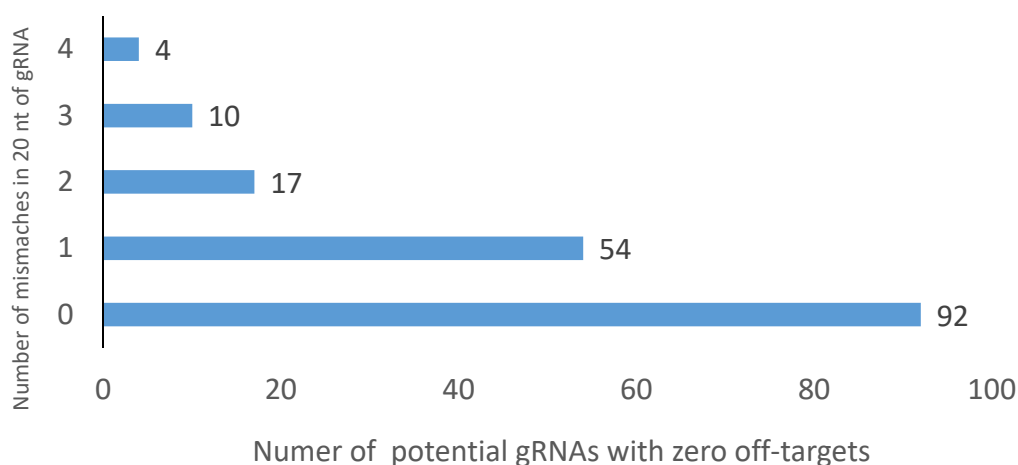
Figure 5.5 Off-target analysis of candidate CRISPR-Cas9 targets in exon 9 of *ATRX*. The results of the short sequence blast against the primary assembly of human genome (Dec. 2013, hg38, GRCh38; using Blast+ v2.5) are shown. It shows the number of potential gRNAs in exon 9 of *ATRX* with zero off-targets that were counted with different number of mismatches in 20 nt of gRNA sequence (zero to 4 mismatches). Mismatches

interpretation; 0 = 20/20 identity, 1 = 19/20 identity, 2= 18/20 identity, 3= 17/20, 4=16/20 identity.
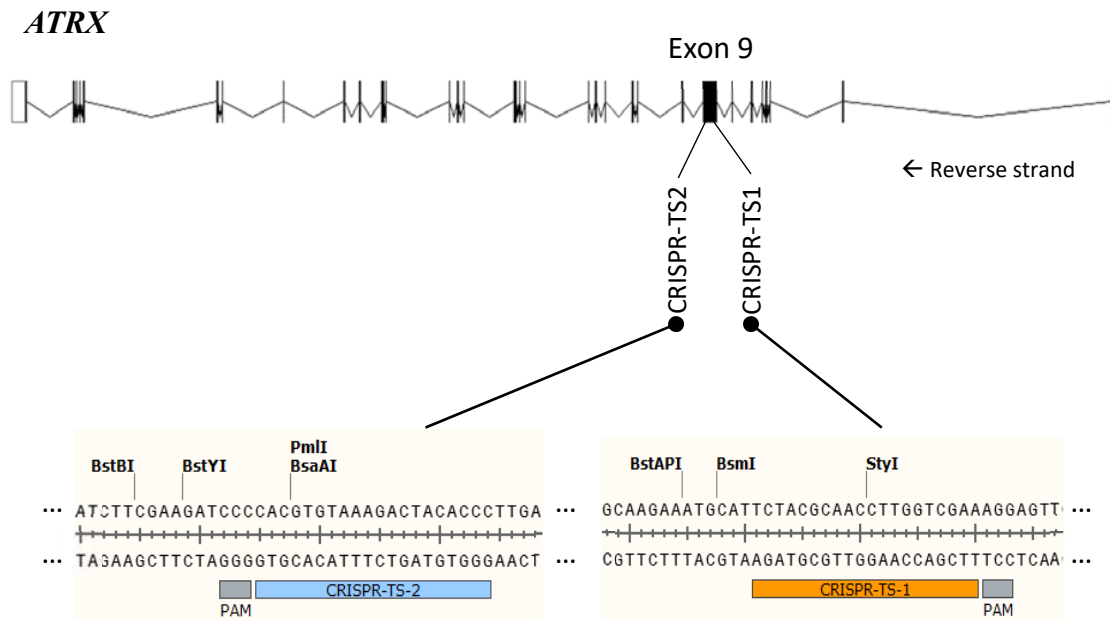
Figure 5.6 Diagram showing the position of the selected CRISPR-target sites in *ATRX* exon 9. *ATRX* gene, the boxes showed 35 exons that ordered from right to left. Below is the location of CRISPR-target sites (CRISPR-TS1 and CRISPR-TS2) in exon 9.

5.3.1.3 Validating the selected CRISPR target sites in TCL1

Prior to cloning the CRISPR-TS-1 and CRISPR-TS-2 in pLEICS-97 and pLEICS-122 (CRISPR vectors of family 95) respectively, the presence of the selected targets in exon 9 of *ATRX* in TCL1 was confirmed by Sanger sequencing of PCR amplicons generated from TCL1 DNA. PCR primers ATRX EX9a.F and ATRX EX9a.R were used to generate the ATRX-EX9a amplicon (695 bp), which contained the CRISPR-TS-1 target site, and ATRX EX9c.F and ATRX EX9c2.R were used to amplify ATRX-Ex9c amplicon (537 bp) containing the CRISPR-TS-2 target site (Appendix 5.1). The amplicons were purified by the Zymoclean gel extraction method and sequencing conducted with ATRX EX9a.R for the ATRX-EX9a amplicon, and with ATRX EX9c.F

for ATRX-EX9a amplicon. The Sanger sequencing was showed that both CRISPR-TS-1 and CRISPR-TS-2 target sequences are present in TCL1 cells (Appendix 5.2).

5.3.1.4 Cloning and validation of CRISPR vectors

In this work, the pLEICS-97 (wild type SpCas9) and pLEICS-122(high fidelity wild type SpCas9) all-in-one vectors of family 95 (provided by the Protein Expression Laboratory (PROTEX) facility, University of Leicester) were selected for cloning of the CRISPR target sequences. These vectors can be used to express wild type SpCas9, the non-coding guide RNA sequence of the selected ATRX-Exon 9 CRISPR target sites (CRISPR-TS-1, CRISPR-TS-2), and the puromycin resistance gene for antibiotic selection. The cloning of the selected CRISPR target sequences was conducted by the PROTEX facility (University of Leicester) using a recombination-based cloning method to insert the U6 promotor-single guide RNA sequence into the selected vector, as described in section 2.2.14 .

A CRISPR-Cas9 called PX459v2.0, which is a version of all-in-one CRISPR-Cas9 editing plasmid share the same features of pLEICS-vector family 95, but with no 20 nt gRNA sequence, was ordered from Addgene to use as an empty vector control. PX459v.2.0 also contains the wild type SpCas9 sequence, a puromycin resistance gene, and a U6-promoter, but without insertion of guide sequence. The size of plasmid construct after cloning is the same for both pLEIC-ATRX-CRISPR-TS-1 and pLEIC-ATRX-CRISPR-TS-2, which is 9183 bp (Figure 5.7 B). To verify the map of plasmid constructs after cloning, restriction digestions were carried out using a double digest with *Eco*RV+ *Not*I, and a triple digest with *Eco*RV, *Not*I and *Pvu*I. The restriction sites and predicted fragment sizes are shown in (Appendix 5.3). Each of these restriction enzymes has a unique cutting site within the plasmids: *Eco*RV in the SpCas9 sequence, *Not*I in the linker sequence near the puromycin resistance gene, *Pvu*I in the ampicillin resistance gene (Figure 5.7 A). The double and triple digests generated the predicted fragments in pLEIC-ATRX-CRISPR-TS-1_C10 and pLEIC-ATRX-CRISPR-TS-2_C11.
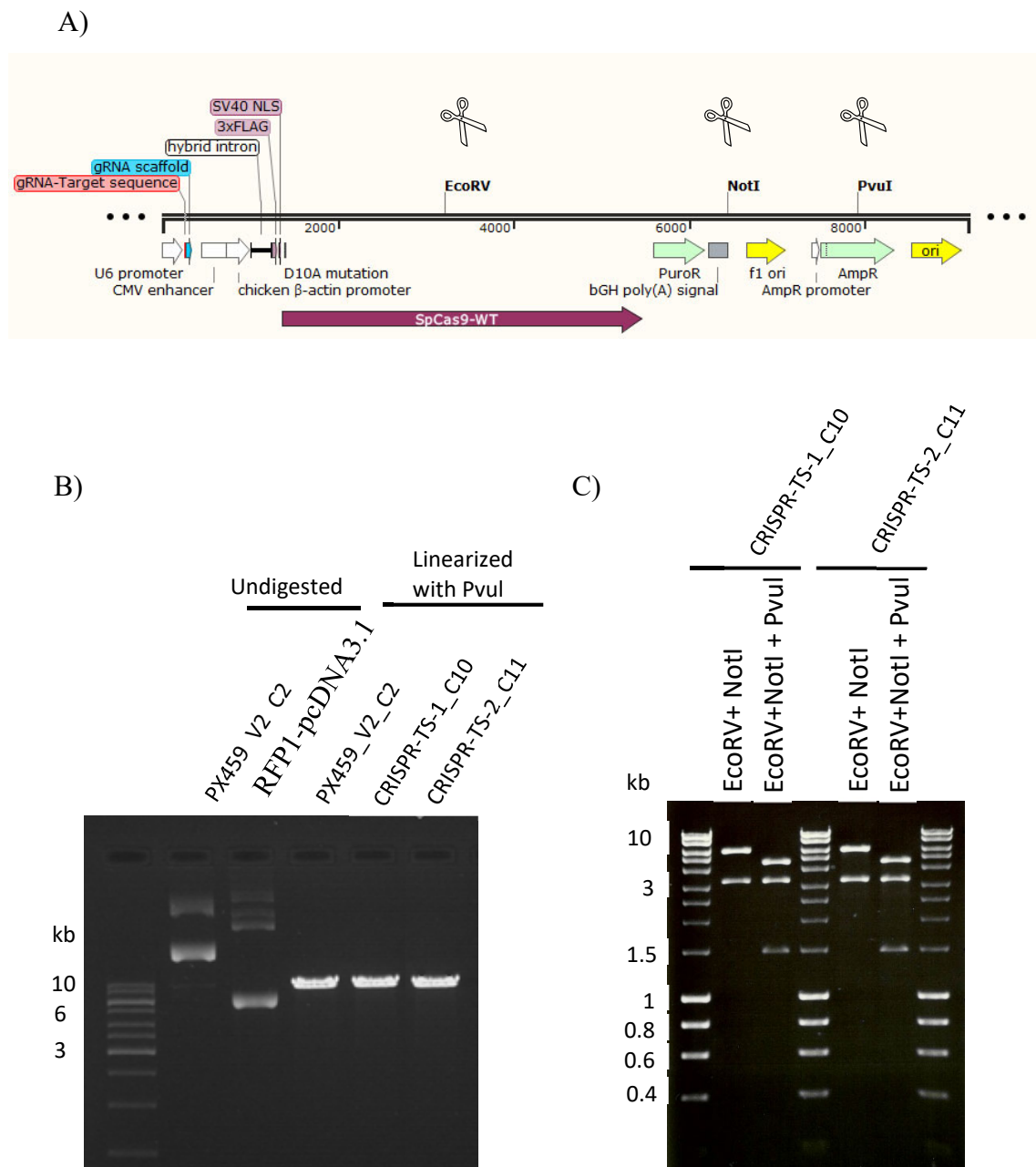
Figure 5.7 Restriction digestion of the CRISPR vectors. (A) Schematic representation of map of the pLEIC-CRISPR vectors after cloning the guide RNA sequence (9183 bp). The unique restriction sites of *Eco*RV, *Not*I, *Pvu*I are shown. (B) Linearization of CRISPR-Cas9 vectors using Pvul. (C) Restriction digest mapping using double digest (EcoRV+ NotI), and triple digest (EcoRV+ NotI + PvuI).

Furthermore, the presence of U6-sgRNA sequence, SpCas9, and PuroR gene were in pLEIC-ATRX-CRISPR-TS-1_C10 and pLEIC-ATRX-CRISPR-TS-2_C11 were confirmed by sequence analysis. A virtual vector sequence was used as a reference in a multiple alignment using MacVector. The alignment showed perfect insertion of a U6 promoter and sgRNA target sequence in both pLEI-CRISPR-TS-1_C10 and pLEI-CRISPR-TS-2_C11 (Figure 5.8). In a similar way, alignment of sequences across Cas9 (including the codon for the essential aspartic acid (G**A**C) residues (Jinek *et al.*, 2012) showed that pLEI-CRISPR-TS-1_C10, pLEI-CRISPR-TS-2_C11 and PX459v2.0 were all wild type for SpCas9 (Figure 5.9). Sequence alignment across the PuroR gene revealed a silent mutation in glycine codon (GGC>GGA) in pLEI-CRISPR-TS-1_C10 and pLEI-CRISPR-TS-2_C11 (Figure 5.10) compared to the PuroR sequence present in PX459v2.0. This silent mutation in pLEI-CRISPR-TS-1_C10 and pLEI-CRISPR-TS-2_C11 is not expected to change the resistance to puromycin following transfection into human cells.
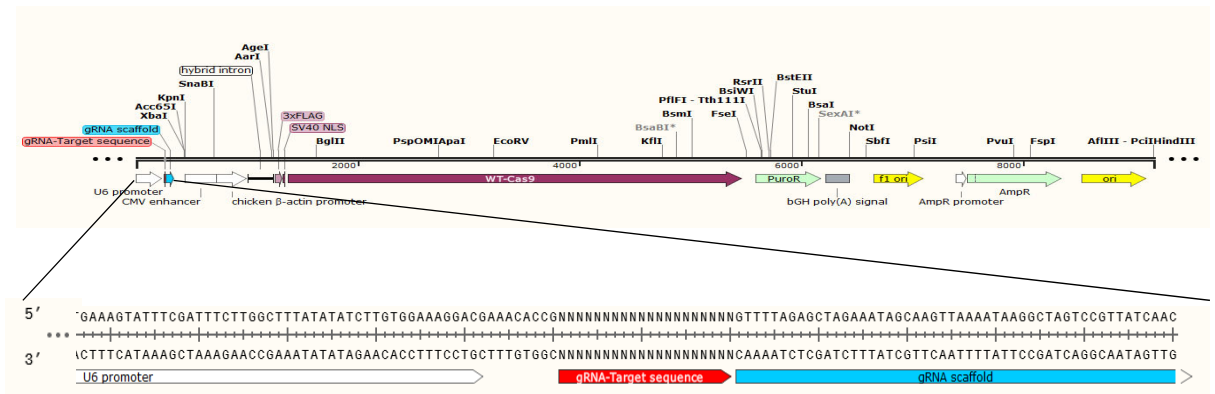
Figure 5.8 Verification of the CRISPR-target sequences by Sanger sequencing. (A) Schematic representation of pLEIC-CRISPR map after cloning the guide sequence (9183 bp). The gRNA-target sequence was replaced with 20 Ns. This sequence was used as a reference for alignment using MacVector. (B) Sequencing alignment of the TS-1 and TS-2 targets sequences from the pLEI-CRISPR-TS-1_C10 and pLEI-CRISPR-TS-2_C11 vectors.
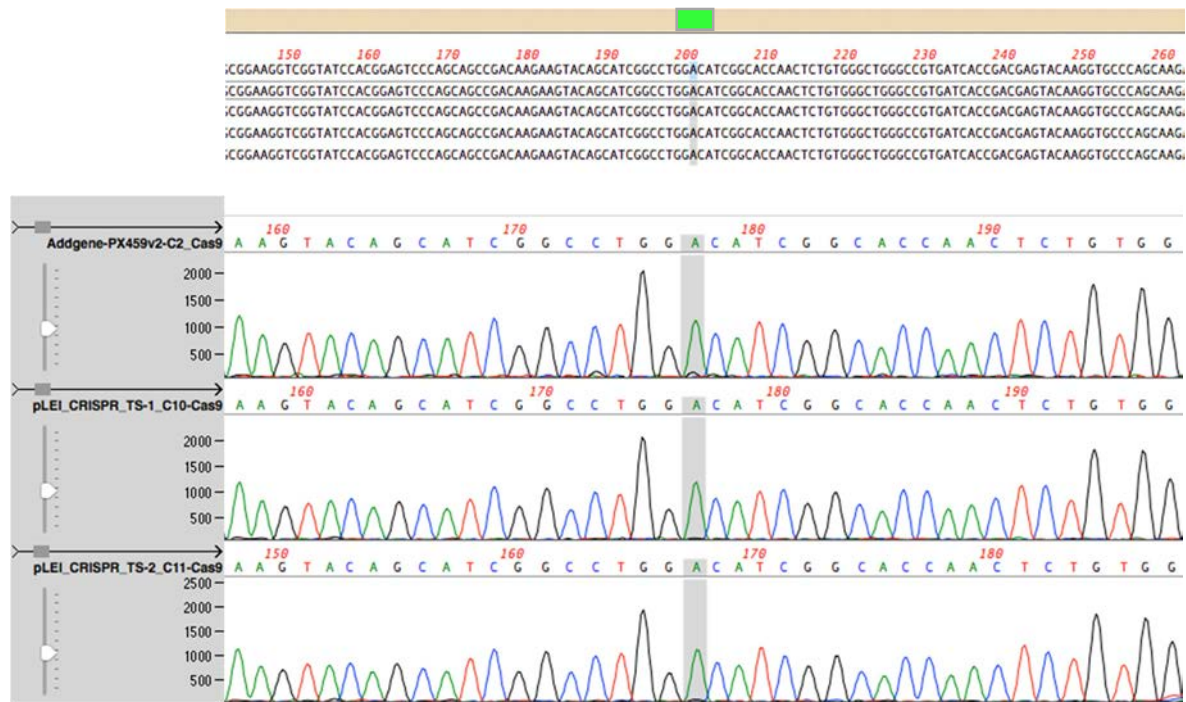
Figure 5.9 Verifcation of the Cas9 sequence in the pLEI-CRISPR-TS-1_C10, pLEI-CRISPR-TS-2_C11and Addgene-PX459v2.0 vectors. (A) Schematic representation of pLEIC-CRISPR map after cloning (9183 bp) with the D10A mutation marked in green. (B) Sequence alignment of part of the Cas9 cassette from the three vectors.

A)

B)
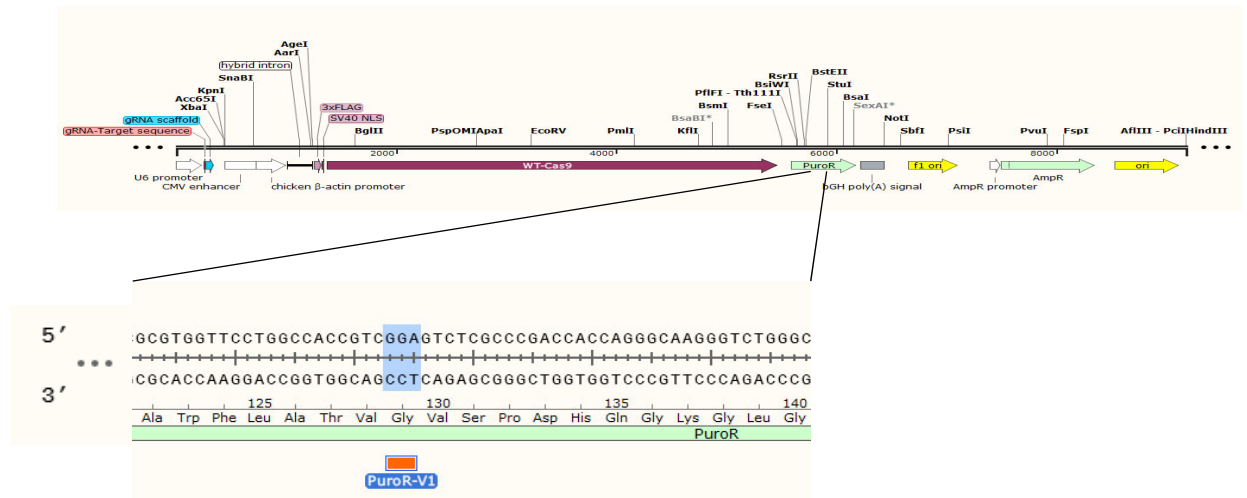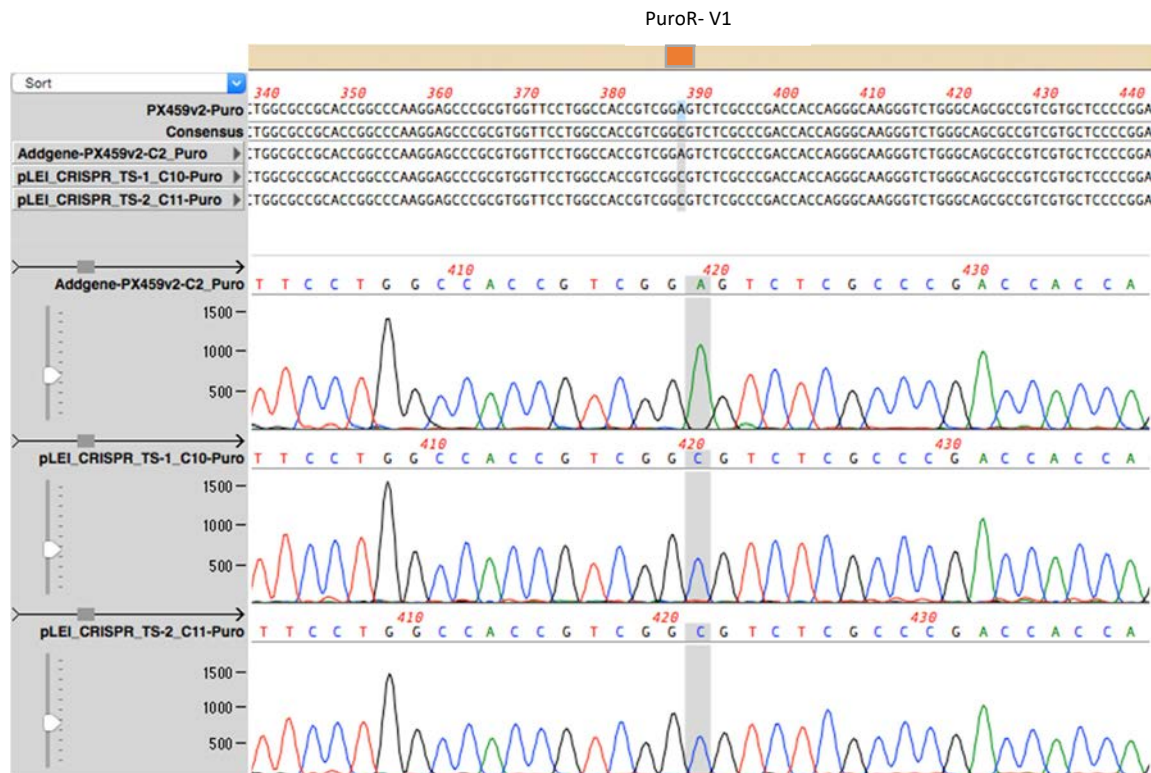
PuroR- V1

Figure 5.10 Verification of the puromycin resistance cassette in the pLEI-CRISPR-TS-1_C10, pLEI-CRISPR-TS-2_C11 and Addgene-PX459v2.0 vectors. (A) Schematic representation of pLEIC-CRISPR map with the PuroR-V1 (variant) marked in orange. (B) Sequence alignment of the PuroR cassette showed the in pLEI-CRISPR-TS-1_C10 and pLEI-CRISPR-TS-2_C11vectors carry a silent A>C substitution in a glycine codon.

5.3.1.5 Optimising CRISPR vector delivery into TCL1 cells

Electroporation was used to deliver the CRISPR-Cas9 expression plasmids into the TCL1 cells using an Amaxa Nucleofector system as described in 2.2.15 Nucleofection into TCL1 cells was optimised using a plasmid RFP1-pcDNA3.1/zeo (a gift from Dr Victoria Cotton), which expresses the monomeric red fluorescent protein 1 (mRFP1) with maximum emission at 607 nm. Nucleofection was conducted in the Ingenio transfection solution (Mirus, UK) and five nucleofection programs for mammalian primary cell lines with were tested to identify the program that gave the highest delivery efficiency into TCL1 cells (Table 5.1 & Figure 5.11). The A-024 program gave the best results with 53% cell survival and 14% plasmid uptake consistent with the findings of Chu et al. (1987).

Table 5.1 Summary of plasmid delivery efficiency into TCL1 cells. Program A-024 showed the best combination of cell survival and plasmid uptake.

Amaxa's Program For mammalian primary cell lines	Total screened cells	% Delivery by RFP	% Survival cells
A-024	207	14	53
U-012	187	4	39
U-023	171	11	33
T-016	118	10	15
V-013	241	5	63

A)
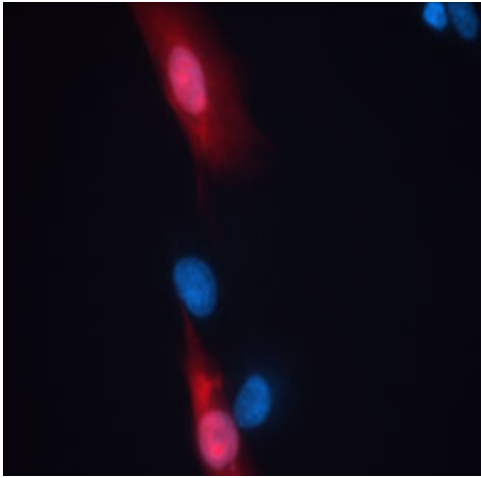
B)
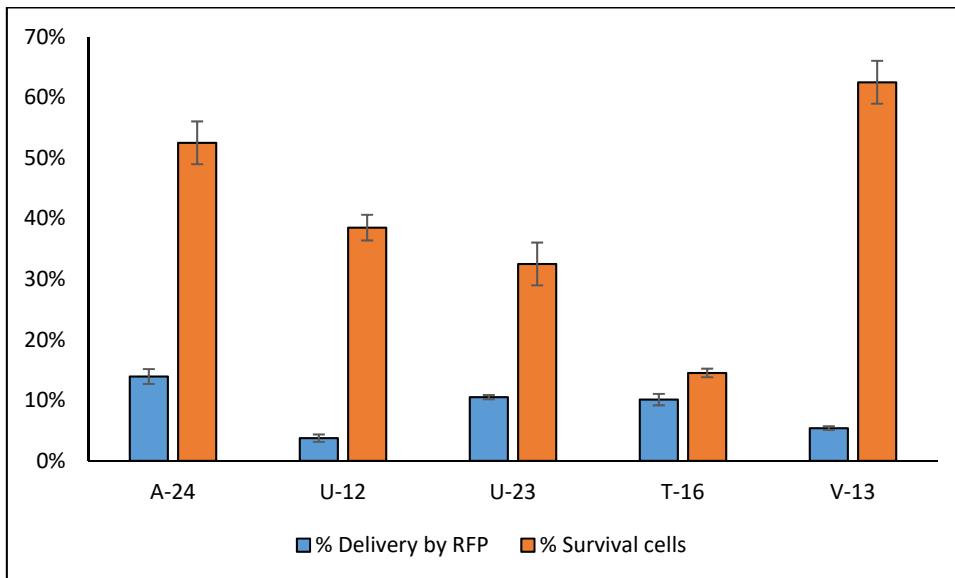
Figure 5.11 Optimising plasmid delivery conditions into TCL1. (A) Confirmed plasmid delivery by direct detection of RFP signal. Cells that express the RFP protein (red) can be counted. The cells were fixed, and the nuclei were stained with DAPI and cells visualised using a fluorescent microscope. (B) Assessment of plasmid delivery efficiency measured as the percentage of cells expressing the RFP reporter and by measuring cell survival using the trypan blue assay. Both delivery and survival measurements were taken 24 hr after transfection with the RFP reporter construct. n= two independent experiments.

5.3.1.6 Transfection efficiency and production of clones that survive crisis

To maximise delivery efficiency of the Addgene-PX459v2.0-Empty_Vector, pLEIC-ATRX-CRISPR-TS-1 and pLEIC-ATRX-CRISPR-TS-2 vectors into TCL1 cells, nucleofection was carried out using the Amaxa A-024 program in parallel transfections with 1×10^6 cells per reaction. Three different combinations were used: transfection with the control Addgene-PX459v2.0 empty vector alone, transfected with pLEIC-ATRX-CRISPR-TS-1, and co-transfection with both pLEIC-ATRX-CRISPR-TS-1 & pLEIC-ATRX-CRISPR-TS-2 in order to improve the editing efficiency at ATRX-Exon 9 locus. Puromycin selection was added in combination with G-418 (G418 was used to maintain SV40-T antigen expression) to the media 48 hours after transfection to allow the CRISPR-Cas9 plasmid to be expressed. Three independent experiments were conducted and approximately 2-3 weeks later, single clones were expanded into 24-well plates using Scienceware® cloning discs (Sigma-Aldrich). Clones that continued to divide were counted and transferred to 6 well-plates.

Table 5.2 shows that similar numbers of clones emerged following transfection with the different vectors but none of the empty vector transfected clones and only six of the clonal populations (CP) with ATRX editing went on to survive crisis. These include two (CP_21, CP_105) of 110 clones that formed following transfection with CRISPR-TS-1 and four (CP_05, CP_15, CP_16, and CP_38) out of 55 clones that formed following co-transfected with CRISPR vectors for both ATRX target sites. The ATRX-targeted and empty vector-transfected clones all reached the crisis in the same time period with most of them stop dividing before reaching confluence in 6 well plates (~0.5-1 $\times 10^6$ cells per well). It was showed previously that the spontaneous immortalisation frequency of human fibroblast SV40-T-antigen-transfected clones is around 3×10^{-7} (Shay et al. 1993) and higher in mammalian epithelial cells (10×10^{-5}). In this work, the frequency of immortalisation among SV40-LT- transfected human fibroblast was 1.35×10^{-7} following transfection with a single ATRX-editing vector (CRISPR-TS-1) and much higher 2.75×10^{-6} following transfection with two ATRX-editing vectors (CRISPR-TS-1 & CRISPR-TS-2).

Table 5.2 Summary of clones generated following transfection and the number that survived crisis.

Transfection	Total Number electroporated cells	Total number of expanded clones	Frequency of surviving clones 5 weeks after transfection	Number of clonal populations that survived crisis
Empty Vector	10 X 10^6	42	33/42 (78.6%)	0
CRISPR-TS-1	27 X 10^6	110	80/110 (72.7%)	2
CRISPR-TS-1 & CRISPR-TS-2	11 X 10^6	55	37/55 (67.3%)	4

5.3.1.7 Growth rate analysis

Prior to the start of the work, the CCD18LU normal primary cell line (lung fibroblast, female) was minimally expanded for prior to transfection with the pSV3neo, SV40-LT expressing vector. Following G418 antibiotic selection, single colonies were picked and expanded to 6 well plate (~19 PDs). The TCL1 clone was then expanded for approximately 13 PDs at which point it was frozen in liquid nitrogen. In this work, the population doublings were counted and growth rate analysis conducted for the TCL1 after thawing the cells. Thus, the TCL1 cells reached crisis ~22 PDs after they were thawed from liquid nitrogen, but the total estimated number of PDs that TCL1 underwent prior to crisis was ~70 PDs (Figure 5.12). Consequently, following expansion and transfection the majority of emerging clones hit crisis before they reached confluence in 6 well plates and the ATRX status remained unknown until the early post-crisis stage.
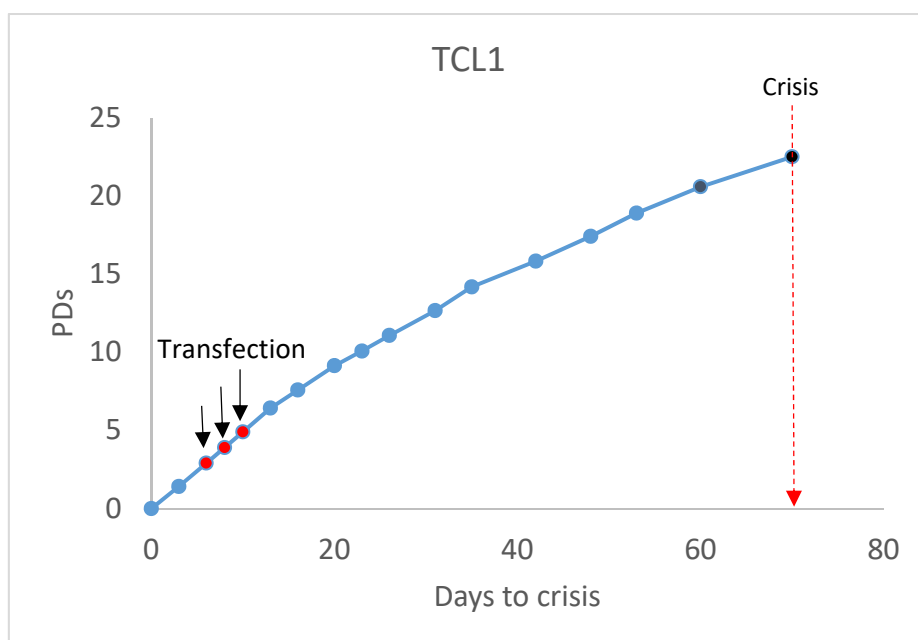
Figure 5.12 Population doubling analysis for TCL1. Three red nodes highlight the three timings of the transfection trials for CRISPR-Cas9 - ATRX editing. It also showed the progressive slowdown in growth rate towards the crisis.

Entry into crisis was characterised by a balanced status of cells continuing to divide and cell death resulting in a decrease in cell survival, while late crisis was characterised by a substantial increase in cell death. To characterise the mechanism by which cells were dying, programmed cell death was investigated in TCL1 by looking for phenotypic markers of apoptosis including chromatin condensation and nuclear fragmentation events identified by DAPI staining (Figure 5.13). Furthermore, the frequency of TCL1 cells undergoing apoptosis was identified by FACS analysis using the Annexin V/Propidium Iodide apoptosis assay as described in 2.2.19. The results showed that ~45% of cells at crisis (approximately PD 22) were undergoing early apoptosis (which activated the caspase, but still had intact cell membrane), and only 8% were in late apoptosis (where cells have activated caspase system and showed loss of membrane integrity). This suggested a delayed induction of apoptosis, possibly as a results of p53 inactivation by SV40-LT, however, the mechanism by which apoptosis is activated in SV40-T-expressing cells is poorly understood.
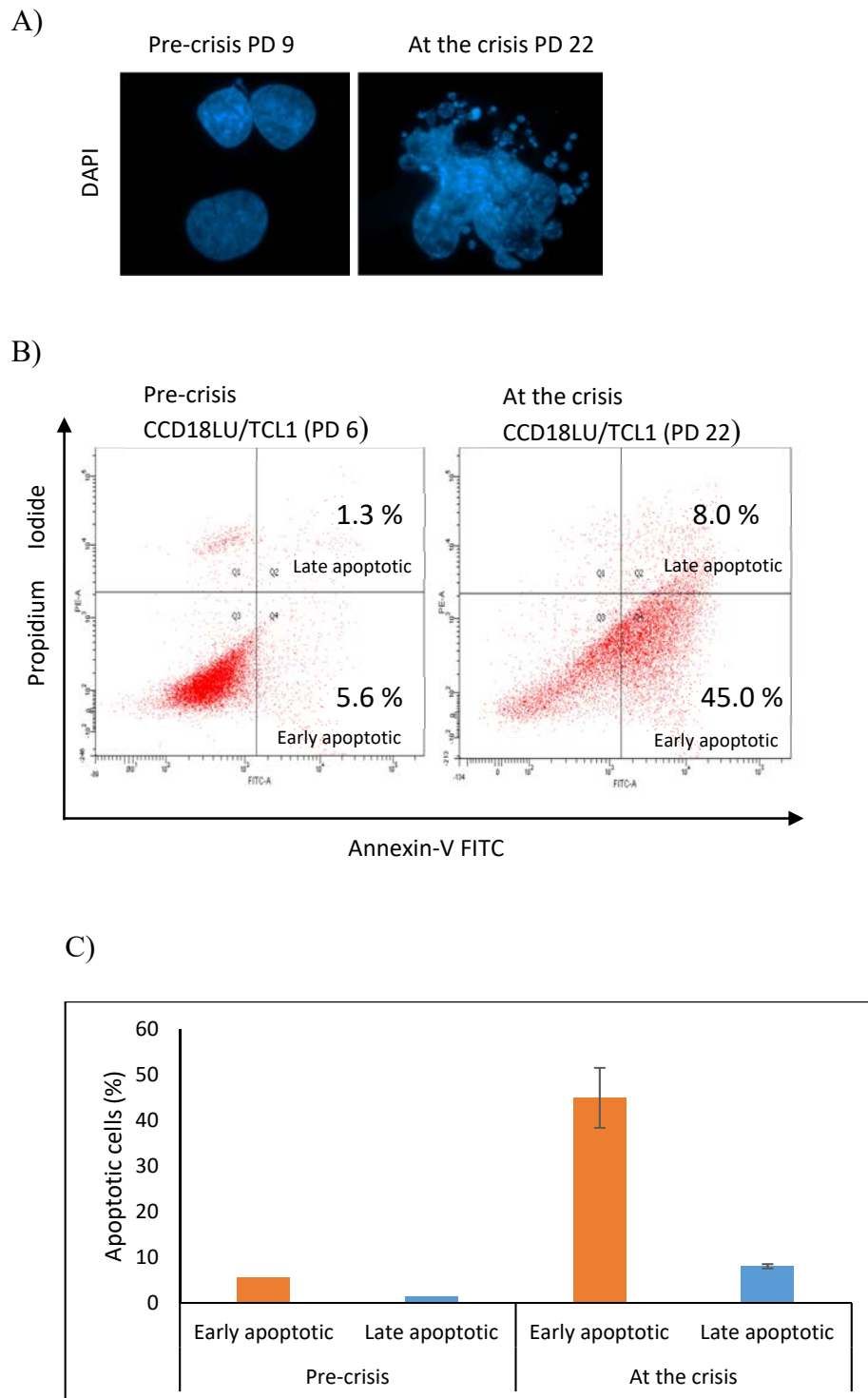
Figure 5.13 Apoptosis mediated cell death in TCL1 at the crisis. (A) DNA fragmentation detected using DAPI staining in TCL1 at PD 21. (B-C) Frequency of early and late apoptosis were measured using FACS analysis of FITC-conjugated-Annexin V and PI staining.

Following transfection with the ATRX-targeting CRISPR/Cas9 vectors, many colonies were expand as indicated above. The six expanded clones that hit and survived crisis passed through another bottleneck with a limited number of subclones (estimated by visual inspection, range 1 -12 subclones) giving rise to the post-crisis immortalised clonal populations (CP) (Table 5.3). Cell counts were obtained for all cultures pre-crisis, and directly after the crisis. Thus, the number of PDs were adjusted according to the estimated number of colonies that gave rise to each post-crisis clonal population. Clonal populations were cultured for approximately 40 PDs after the crisis and cell pellets collected at various time points. Cells collected at two time points (T1 and T2, Figure 5.14) were used for multiple downstream analyses.

Table 5.3 Estimated number of colonies emerging from crisis for each clonal population.

Clonal Population ID	Estimated duration of crisis (Days)	Estimated No. of subclones emerging from crisis	Estimated No. of PDs post-crisis to reach $1x10^6$ cells
CP_05	59	7	~17 PDs
CP_15	63	12	~ 16.3 PDs
CP_16	50	8	~ 18 PDs
CP_21	50	4	~17 PDs
CP_38	50	1	~20 PDs
CP_105	50	2	~ 19 PDs
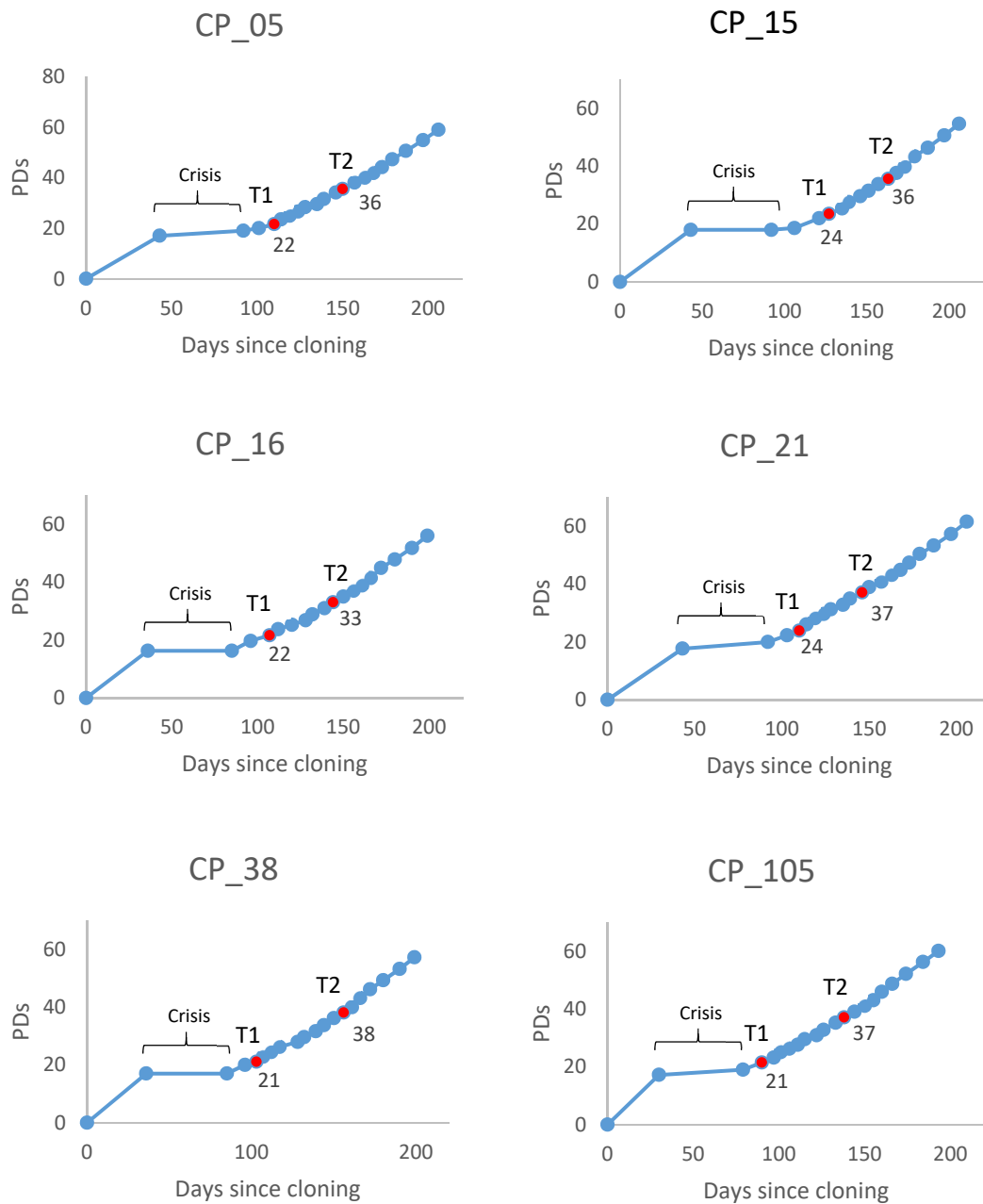
Figure 5.14 Growth rate analysis of clonal populations that survived crisis. The growth curve start (Day zero) when each clone was picked following transfection with the ATRX-CRISPR/Cas9 targeting vector(s). Each clone entered crisis at a similar time and crisis lasted was between 50-60 days. The red dots identify the early (T1) and late (T2) post crisis time points used of gene expression analysis. PDs = Population doublings

5.3.2 Assaying for *ATRX* knockout in post-crisis clonal populations

Initially, *ATRX* knockout was assessed by looking for loss of ATRX protein expression using IF staining (primary antibody ATRX- H-300 detected by the Alex488 dye-Anti-IgG secondary antibody). The post-crisis clonal populations at T1 were assessed along-side the empty vector transfected TCL1 cells before and at the crisis. The results showed complete loss of ATRX in all the post crisis CPs, but not in pre-crisis or at the crisis (Figure 5.15 A). These findings were confirmed by western blot detection analysis that showed the presence of full length ATRX in TCL1-Empty-Vector-transfected cells, but not in any of post-crisis CPs at time point 2 (T2). HT1080 (TEL+) and U2OS (ALT+) were used as positive and negative controls respectively and β-Tubulin was used as a protein loading control (Figure 5.15 B)
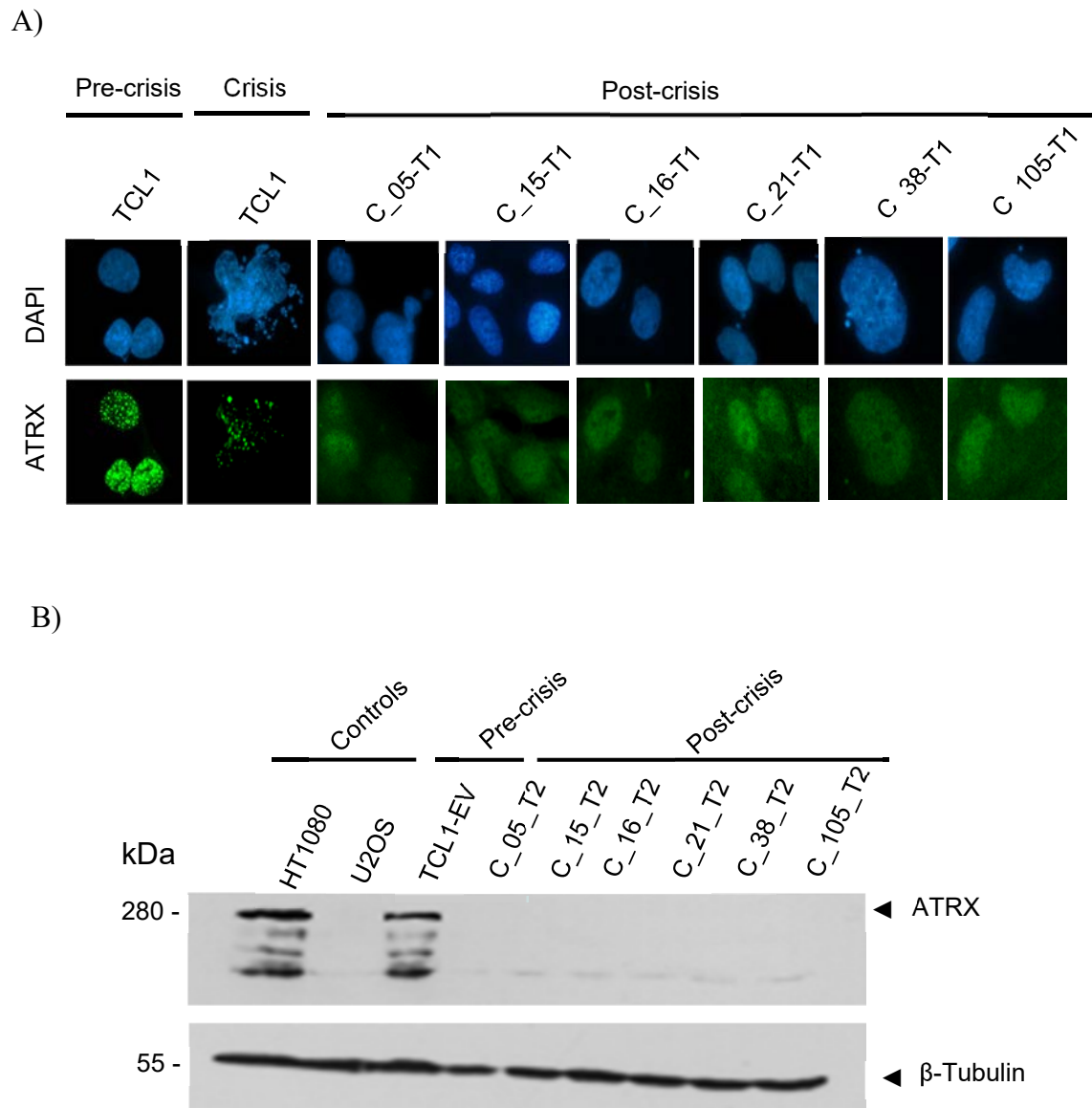
Figure 5.15 ATRX expression was not detected in post-crisis CPs. (A) IF staining to detect ATRX in interphase cells counterstained with DAPI (blue). Pre-crisis TCL1cells were used as controls. (B) Western blot detection of ATRX. HT1080, U2OS and TCL1 empty vector pre-crisis cells were used as assay controls. β-Tubulin was used as loading control.

To characterise the mutations introduced by CRISPR-Cas9 induced NHEJ repair of DSB at two positions in *ATRX* exon 9, two amplicons *ATRX*-Ex9a and *ATRX*-Ex9c that contained the CRISPR-TS-1 and CRISPR-TS-2 targets respectively were generated. Size separation by agarose gel electrophoresis of the *ATRX*-Ex9a and *ATRX*-Ex9c amplicons showed two different deletions in *ATRX*-Ex9a in CP_38 and deletions in *ATRX*-Ex9c in CP_05 and CP_38 (Figure 5.16).
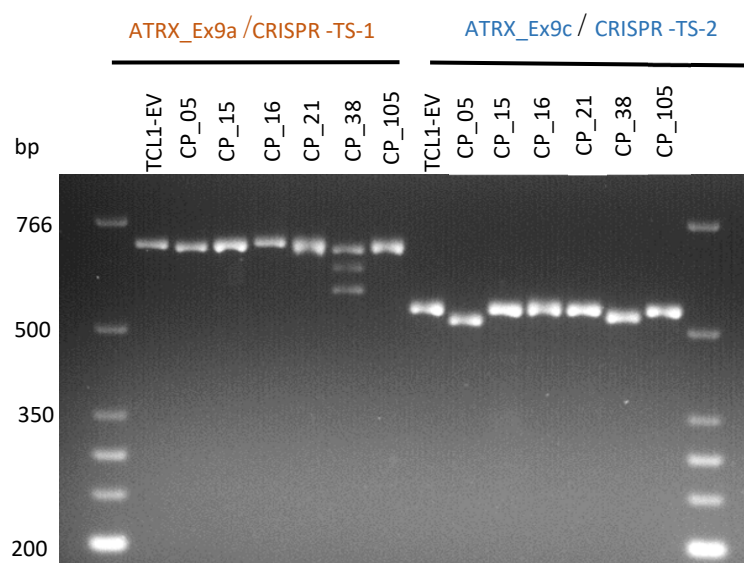
Figure 5.16 Identification of deletions within *ATRX*-Exon 9. Gel electrophoresis of the undigested ATRX-Ex9a and ATRX-Ex9c amplicons showed two different deletions in the ATRX-Ex9a amplicon in CP_38 as well as a full-length product and deletions in ATRX-Ex9c in CP_05 and CP_38.

To investigate other possible mutations the T7 endonuclease I (T7EI) assay was carried out as on DNA from the various CPs as described in 2.2.2.12. The T7 endonuclease I recognises and cleaves DNA heteroduplex formed by single base mismatches or indel loops created by mismatched between several bases (Mashal et al. 1995; Vouillot et al. 2015). A DNA sample that is heterozygous for a wild type and mutated copy of the amplified target sequence form homo- and hetero-duplexes following denaturation and re-annealing but only the heteroduplex containing products are cleaved by T7 endonuclease I . In addition the fraction of cleaved amplicons is also dependent on the efficiency of the T7 endonuclease I enzyme to recognise and cut the heteroduplex. In this work, the T7EI assay showed that *ATRX* had been edited at the CRISPR-TS-1 target in in CP_21, CP_38, and CP_105 and at the CRISPR-TS-2 target in CP_15, CP_15, and CP_38 (Figure 5.17). Interestingly, the T7 endonuclease I failed to recognise the deletion mutations in the in CP_05. The findings were also confirmed by Sanger sequencing (Appendix 5.4).

A)
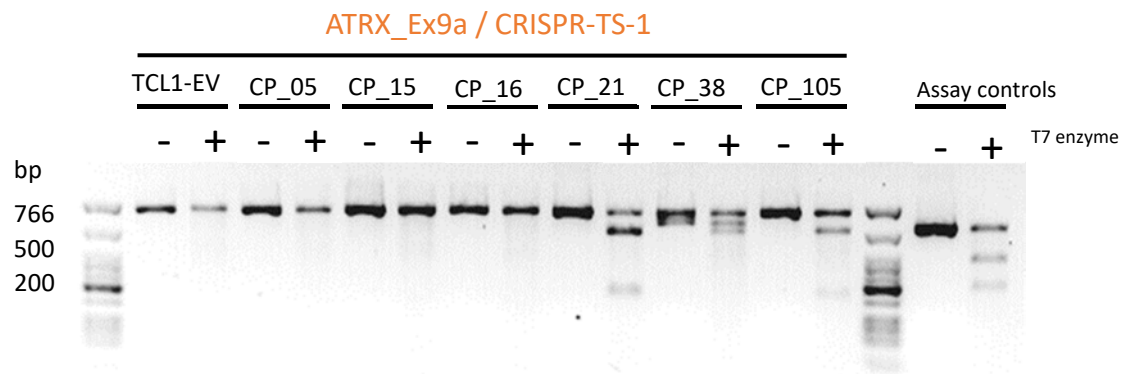
B)
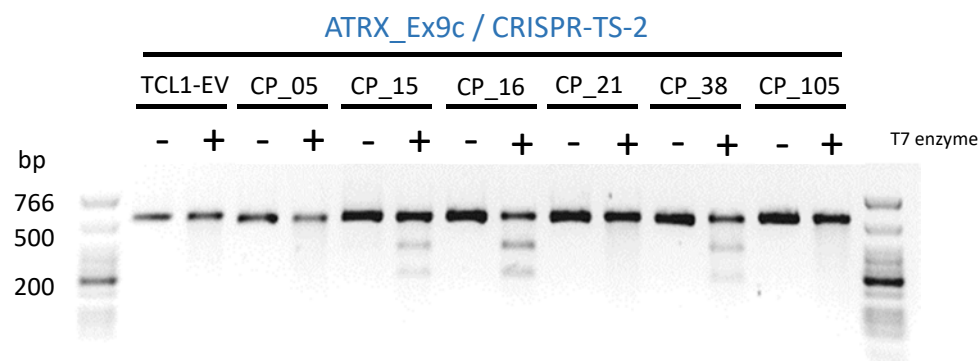
Figure 5.17 Detection of *ATRX* exon 9 editing in post-crisis CPs using the T7 endonuclease I assay. (A) Screening CPs for editing at the CRISPR-TS1 target in the ATRX_Ex9a amplicon. (B) Screening CPs for editing by CRISPR-TS2 target in the ATRX_Ex9c amplicon. DNA from pre-crisis SV40-TCL1-EV was used as a negative control for CRISPR editing at both targets.

Further confirmation of the indel mutations that arose from *ATRX* editing by CRISPR-Cas9 in the post-crisis CPs was identified from the whole exome sequencing (WES) analysis that will be described fully in Chapter 6. The Mutect2 somatic caller was used to collect to generate list of indels (described in the methods). Inspection of the identified indels was carried out by visualising the .bam files in the Integrative Genomics Viewer (IGV). IGV also was used to calculate the frequency of mutated reads at each of CRISPR-TS-1and CRISPR-TS-2. It showed that the frequency of reads with Indel was between 46 to 86% at CRISPR-TS-1, and 46 to 100% at CRISPR-TS-2 (Table 5.4). As 100% of the reads that showed a mutation at CRISPR-TS-2 target in CP_05 and CP_38 it was deduced that both alleles had been mutated. Whereas, when approximately 50% of reads shared the same mutation it was assumed that only one allele carried that particular mutation. The presence of 82% reads with a mutation at CRISPR-TS-1 in CP_21 and 60% at CRISPR-TS-1 in CP_16 might be explained by clonal variation. Most of the *ATRX* mutations were deletions that varied in size from 1 to 26 base pair. However, CP_16 showed an insertion of a cytosine (C) in 36% of reads at X:76938774, and a two base deletion (AC) in 60% of reads at X:76938772.

To understand the consequence of these indel mutations on *ATRX* expression, missense prediction analysis was conducted using SIFT 4G annotator (Vaser *et al.*, 2016). The SIFT 4G analysis showed that the detected mutations are likely to disrupt the three *ATRX* transcripts. Two of the transcripts (ENST00000373344 and ENST00000395603) code for full length or nearly full length ATRX protein (ATRX-201 (2492aa), and ATRX-202 (2454aa)), while the third ENST00000400862 codes for an ATRX isoform that contains the ADD but not the helicase domains. All the identified *ATRX* exon 9 mutations are predicated to cause Nonsense Mediated Decay (NMD) of the transcripts and no production of a functional ATRX protein (Table 5.4). This prediction is supported by western blot results at time point 2 (T2) (Figure 5.15 B).

Table 5.4 All the *ATRX* mutations identified in the post-crisis CPs are predicted to result in Nonsense Mediated Decay (NMD) of transcripts. The predictions were generated using the *ATRX* indels coordinates and the SIFT 4G annotator (Vaser *et al.*, 2016).

CP_ID	CRISPR-TS	Reads edited %	Coordinates	Ref	Alt	Gene Name	Transcript ID	Substitution Type	AA position change	Effect	Amino acid change	Causes NMD
CP_05	2	16/16 (100%)	X:76939994-76940001	TAGTCTTTACACGTGGGGATCTT	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	249-2492	damaging	LRNLGrkelstimdennqwy...(+2228 amino acids)...*->LRNLGvvhnng*	YES
CP_05	2	16/16 (100%)	X:76939994-76940001	TAGTCTTTACACGTGGGGATCTT	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	211-2454	damaging	LRNLGrkelstimdennqwy...(+2228 amino acids)...*->LRNLGvvhnng*	YES
CP_05	2	16/16 (100%)	X:76939994-76940001	TAGTCTTTACACGTGGGGATCTT	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	205-1345	damaging	LRNLGrkelstimdennqwy...(+1125 amino acids)...*->LRNLGvvhnng*	YES
CP_05	1	16/32(50%)	X:76938764-76938786	CCTTTCGA	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	654-2492	damaging	ESDLRrsprvkttplrrpte...(+1823 amino acids)...*->ESDLRhp*	YES
CP_05	1	16/32(50%)	X:76938764-76938786	CCTTTCGA	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	616-2454	damaging	ESDLRrsprvkttplrrpte...(+1823 amino acids)...*->ESDLRhp*	YES
CP_05	1	16/32(50%)	X:76938764-76938786	CCTTTCGA	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	581-1345	damaging	ESDLRrsprvkttplrrpte...(+749 amino acids)...*->ESDLRhp*	YES
CP_15	2	19/41(46%)	X:76938776-76938776	G	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	657-2492	damaging	LRRSPrvkttplrrptetnp...(+1820 amino acids)...*->LRRSPv*	YES
CP_15	2	19/41(46%)	X:76938776-76938776	G	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	619-2454	damaging	LRRSPrvkttplrrptetnp...(+1820 amino acids)...*->LRRSPv*	YES
CP_15	2	19/41(46%)	X:76938776-76938776	G	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	584-1345	damaging	LRRSPrvkttplrrptetnp...(+746 amino acids)...*->LRRSPv*	YES
CP_16	2	9/25(36%)	X:76938774-76938774	-	C	ATRX	ENST00000373344	FRAMESHIFT Inserstion	658-2492	damaging	RRSPRvkttplrrptetnpv...(+1819 amino acids)...*->RRSPRckdytleatdrn*	YES

CP_16	2	9/25(36%)	X:76938774-76938774	-	C	ATRX	ENST00000395603	FRAMESHIFT Inserstion	620-2454	damaging	RRSPRvkttplrrptetnpv...(+1819 amino acids)...*->RRSPRckdytleatdrn*	YES
CP_16	2	9/25(36%)	X:76938774-76938774	-	C	ATRX	ENST00000400862	FRAMESHIFT Inserstion	585-1345	damaging	RRSPRvkttplrrptetnpv...(+745 amino acids)...*->RRSPRckdytleatdrn*	YES
CP_16	2	15/25(60%)	X:76938772-76938773	AC	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	658-2492	damaging	RRSPRvkttplrrptetnpv...(+1819 amino acids)...*->RRSPRkdytleatdrn*	YES
CP_16	2	15/25(60%)	X:76938772-76938773	AC	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	620-2454	damaging	RRSPRvkttplrrptetnpv...(+1819 amino acids)...*->RRSPRkdytleatdrn*	YES
CP_16	2	15/25(60%)	X:76938772-76938773	AC	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	585-1345	damaging	RRSPRvkttplrrptetnpv...(+745 amino acids)...*->RRSPRkdytleatdrn*	YES
CP_21	1	23/28(82%)	X:76939987-76939999	GACAACTCCTTTC	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	249-2492	damaging	LRNLGrkelstimdennqwy...(+2228 amino acids)...*->LRNLGpq*	YES
CP_21	1	23/28(82%)	X:76939987-76939999	GACAACTCCTTTC	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	211-2454	damaging	LRNLGrkelstimdennqwy...(+2228 amino acids)...*->LRNLGpq*	YES
CP_21	1	23/28(82%)	X:76939987-76939999	GACAACTCCTTTC	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	205-1345	damaging	LRNLGrkelstimdennqwy...(+1125 amino acids)...*->LRNLGpq*	YES
CP_21	1	23/28(82%)	X:76939986-76940005	GGACAACTCCTTTCGACCAA	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	247-2492	damaging	CILRNlgrkelstimdennq...(+2230 amino acids)...*->CILRNhnng*	YES
CP_21	1	23/28(82%)	X:76939986-76940005	GGACAACTCCTTTCGACCAA	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	209-2454	damaging	CILRNlgrkelstimdennq...(+2230 amino acids)...*->CILRNhnng*	YES

CP_21	1	23/28(82%)	X:76939986-76940005	GGACAACTCCTTTCGACCAA	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	203-1345	damaging	CILRNlgrkelstimdennq...(+1127 amino acids)...*->CILRNhnng*	YES
CP_21	1	23/28(82%)	X:76939991-76940003	ACTCCTTTCGACC	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	248-2492	damaging	ILRNLgrkelstimdennqw...(+2229 amino acids)...*->ILRNLcpq*	YES
CP_21	1	23/28(82%)	X:76939991-76940003	ACTCCTTTCGACC	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	210-2454	damaging	ILRNLgrkelstimdennqw...(+2229 amino acids)...*->ILRNLcpq*	YES
CP_21	1	23/28(82%)	X:76939991-76940003	ACTCCTTTCGACC	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	204-1345	damaging	ILRNLgrkelstimdennqw...(+1126 amino acids)...*->ILRNLcpq*	YES
CP_38	2	47/47(100%)	X:76938767-76938782	TCTTTACACGTGGGGA	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	655-2492	damaging	SDLRRsprvkttplrrptet...(+1822 amino acids)...*->SDLRRlhp*	YES
CP_38	2	47/47(100%)	X:76938767-76938782	TCTTTACACGTGGGGA	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	617-2454	damaging	SDLRRsprvkttplrrptet...(+1822 amino acids)...*->SDLRRlhp*	YES
CP_38	2	47/47(100%)	X:76938767-76938782	TCTTTACACGTGGGGA	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	582-1345	damaging	SDLRRsprvkttplrrptet...(+748 amino acids)...*->SDLRRlhp*	YES
CP_38	1	19/34(55%)	X:76939989-76940002	CAACTCCTTTCGAC	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	248-2492	damaging	ILRNLgrkelstimdennqw...(+2229 amino acids)...*->ILRNLvhnng*	YES
CP_38	1	19/34(55%)	X:76939989-76940002	CAACTCCTTTCGAC	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	210-2454	damaging	ILRNLgrkelstimdennqw...(+2229 amino acids)...*->ILRNLvhnng*	YES
CP_38	1	19/34(55%)	X:76939989-76940002	CAACTCCTTTCGAC	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	204-1345	damaging	ILRNLgrkelstimdennqw...(+1126 amino acids)...*->ILRNLvhnng*	YES
CP_105	1	19/41(46%)	X:76938776-76938776	GACCAAGGTTG	-	ATRX	ENST00000373344	FRAMESHIFT Deletion	657-2492	damaging	LRRSPrvkttplrrptetnp...(+1820 amino acids)...*->LRRSPv*	YES
CP_105	1	19/41(46%)	X:76938776-76938776	GACCAAGGTTG	-	ATRX	ENST00000395603	FRAMESHIFT Deletion	619-2454	damaging	LRRSPrvkttplrrptetnp...(+1820 amino acids)...*->LRRSPv*	YES
CP_105	1	19/41(46%)	X:76938776-76938776	GACCAAGGTTG	-	ATRX	ENST00000400862	FRAMESHIFT Deletion	584-1345	damaging	LRRSPrvkttplrrptetnp...(+746 amino acids)...*->LRRSPv*	YES

5.3.3 Characterisation of TMM in the post-crisis clonal populations

To characterise the telomere maintenance mechanism in ATRX$^{-/-}$ post-crisis CPs, three markers of ALT activity (APBs, C-circles and telomere length analysis) and TERT transcription (using RNA-seq) were investigated. First, APBs were assayed as described by Henson et al. (2005). Visualisation of PML by IF was achieved with a secondary antibody conjugated to Alexa488 (green) and this was combined with telomere-PNA hybridisation (Cy3 (red) labelled PNA) to detect telomeres in interphase nuclei. Images were captured using ScannR imaging system and colocalisation analysis was conducted using ImageJ software. The percentage of interphase nuclei with APBs ranged from 0.68 to 4.55% and the number of APBs per nucleus was between 3 and 6 (Table 5.5 and Figure 5.18). This suggests that all the post-crisis CPs have activated ALT.

Table 5.5 APBs analysis in *ATRX*-/- at T1 in post-crisis CPs.

ID	Total No. of Nuclei screened	% of nuclei with APBs	Average No. APBs per nucleus that containing APBs
Pre-crisis TCL1	4062	0.00	0
CP_05	1467	3.00	4.65
CP_15	2194	4.20	6.15
CP_16	1626	4.55	6.45
CP_21	2574	3.57	5.51
CP_38	2656	2.64	4.59
CP_105	2190	0.68	3.00
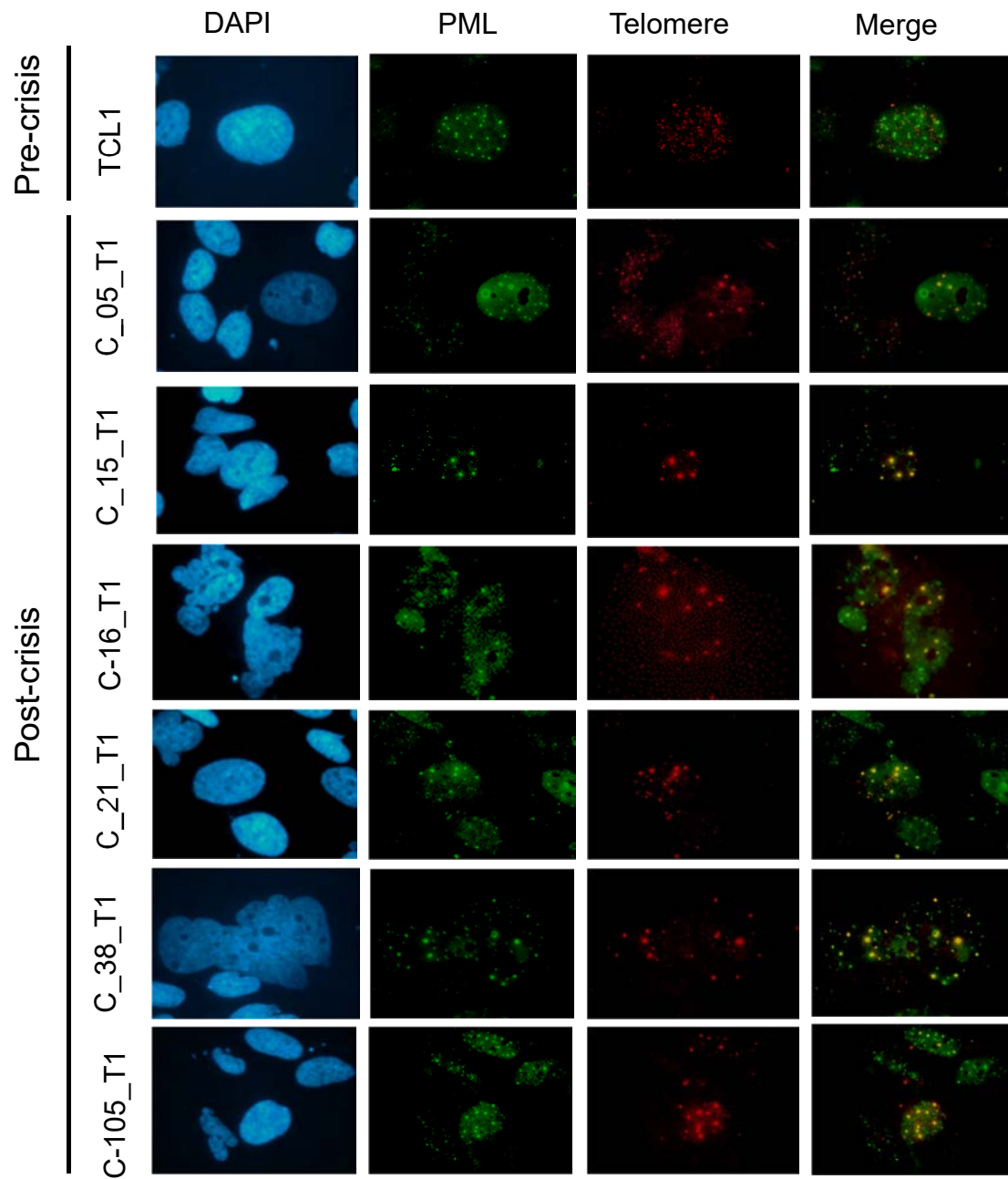
Figure 5.18 ABPs were detected in all *ATRX*-/- post-crisis CPs at T1. IF staining for PML by Alexa 488 (green) at interphase co-localised with Cy3–(CCCTAA)$_3$ telomere PNA probe (red) at interphase, and counterstained with DAPI (blue).

Second, C-circle analysis was conducted using DNA from U2OS (ALT+) cells as positive control and HT1080 (Telomerase+) as negative control. The relative C-circle abundance was obtained by normalising the signal of the telomeric probe (native) to signal of the Alu probe (denatured), then normalising the relative abundance to U2OS using ImageQuant software. Triplicate dot blot analysis was conducted on DNA extracts from the CPs at the late post-crisis time-point (T2). This showed variation in abundance of C-circles relative to U2OS among the post-crisis CPs, ranging between 0.44 - 1.4 (Figure 5.19). The detection of C-circles gave a strong evidence that all the post-crisis CPs had activated the ALT mechanism.

Third, to confirm the activation of ALT in post-crisis CPs, telomere length analysis was carried out using Terminal Restriction Fragments (TRF) analysis as described in 2.2.2.10. Following Southern blot hybridisation, the detected sequences include all telomeres, extrachromosomal linear telomeric repeats and extrachromosomal circular telomeric repeats in ALT+ cells. The TRF analysis showed that all the ATRX-/- post-crisis CPs had highly heterogeneous length telomeres with smears and bands >23 kb. In contrast, the pre-crisis TCL1 cells had TRF smears <10 kb (Figure 5.20). In summary, the three assays used to assess the TMM indicate that all ATRX-/- post-crisis CPs had activated the ALT mechanism.

A)
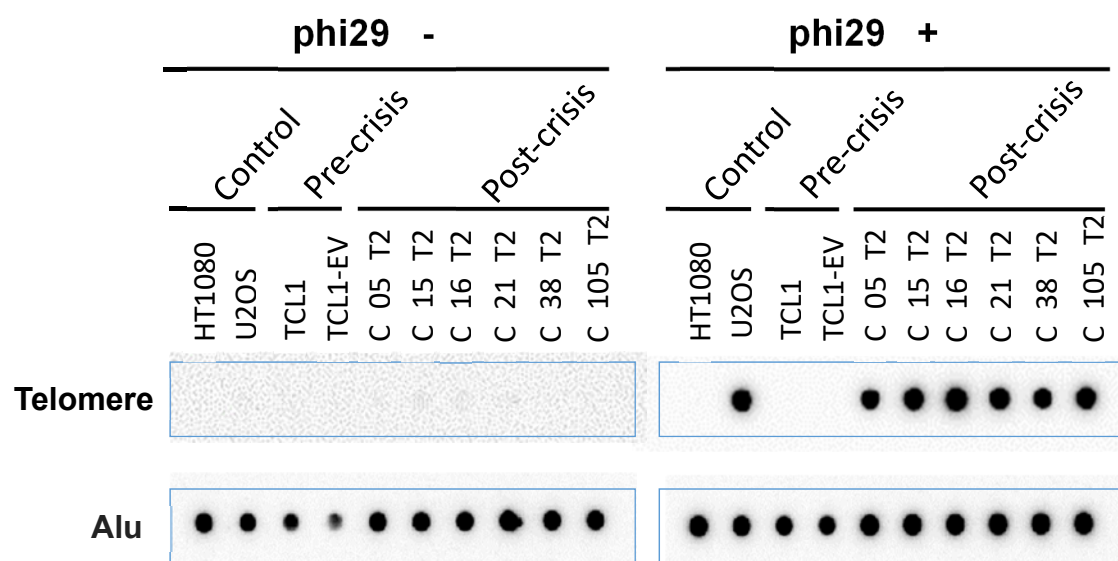
B)
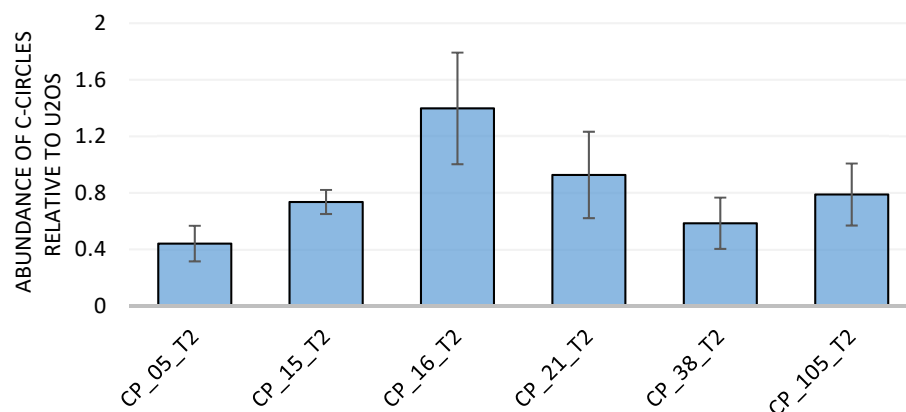
Figure 5.19 C-circles were detected in all ATRX-/- post-crisis CPs. (A) Dot plot of rolling amplification products with/without the phi29 enzyme. Tel= native hybridisation of the telomere probe to the single strand products generated by rolling circle amplification. Alu= denaturing hybridisation to the Alu probe was used to assess input DNA. (B) The bar chart shows the mean of normalised C-circles level ± standard deviation, n = 3 independent experiments. HT1080 and U2OS were used as assay controls.
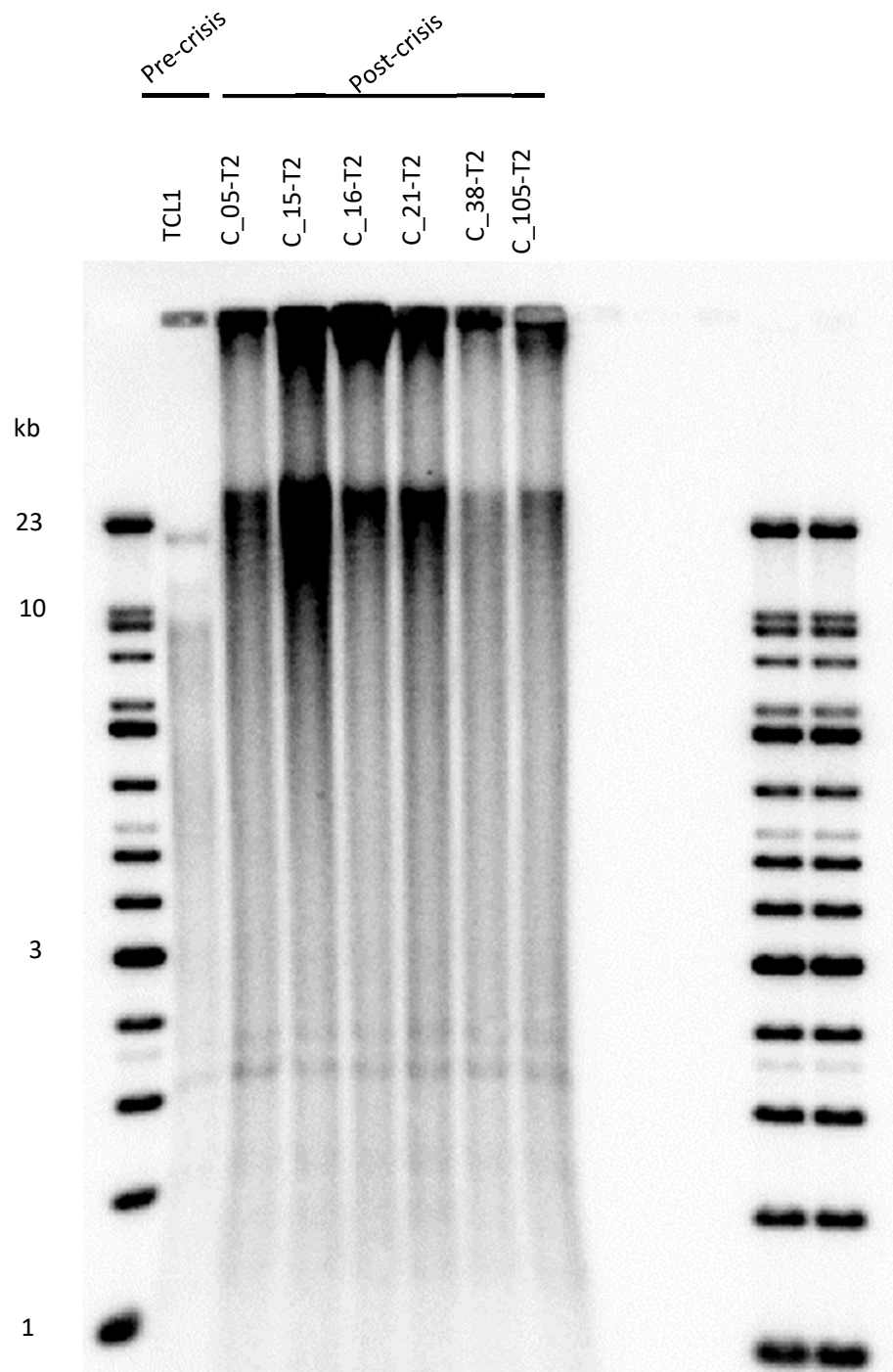
Figure 5.20 Heterogeneous telomere lengths were detected in all ATRX-/- post-crisis CPs. Approximately 6 μg of genomic DNA was digested with *Rsa*I and *Hinf*I, and resolved in a 0.6% agarose gel. The resulting Southern blot was hybridised to a [32]P dCTP labelled telomere probe. The pre-crisis TCL1 had telomere smears <10 kb, but ATRX-/- post-crisis showed heterogeneous telomere length.

To investigate the possibility of telomerase activation in pre-crisis TCL1 or the ATRX-/- post-crisis CPs, the expression of TERT was examined at early (T1) and late post crisis (T2) time points using the RNA-seq data. The .bam files aligned reads were inspected using IGV. This showed no expression of TERT in the pre-crisis TCL1, or in ATRX-/- post crisis CPs (Figure 5.21). Furthermore, the possibility of a mutation in TERT was investigated in pre-crisis TCL1 and ALT+ post-crisis CPs by using aligned reads in .bam files from the WES data from the early post-crisis (T1). The normal CCD18LU was used as reference in IGV. This showed presence of wild type TERT in pre-crisis TCL1, pre-crisis TCL1-EV and all ALT+ post-crisis clonal populations (Figure 5.22).
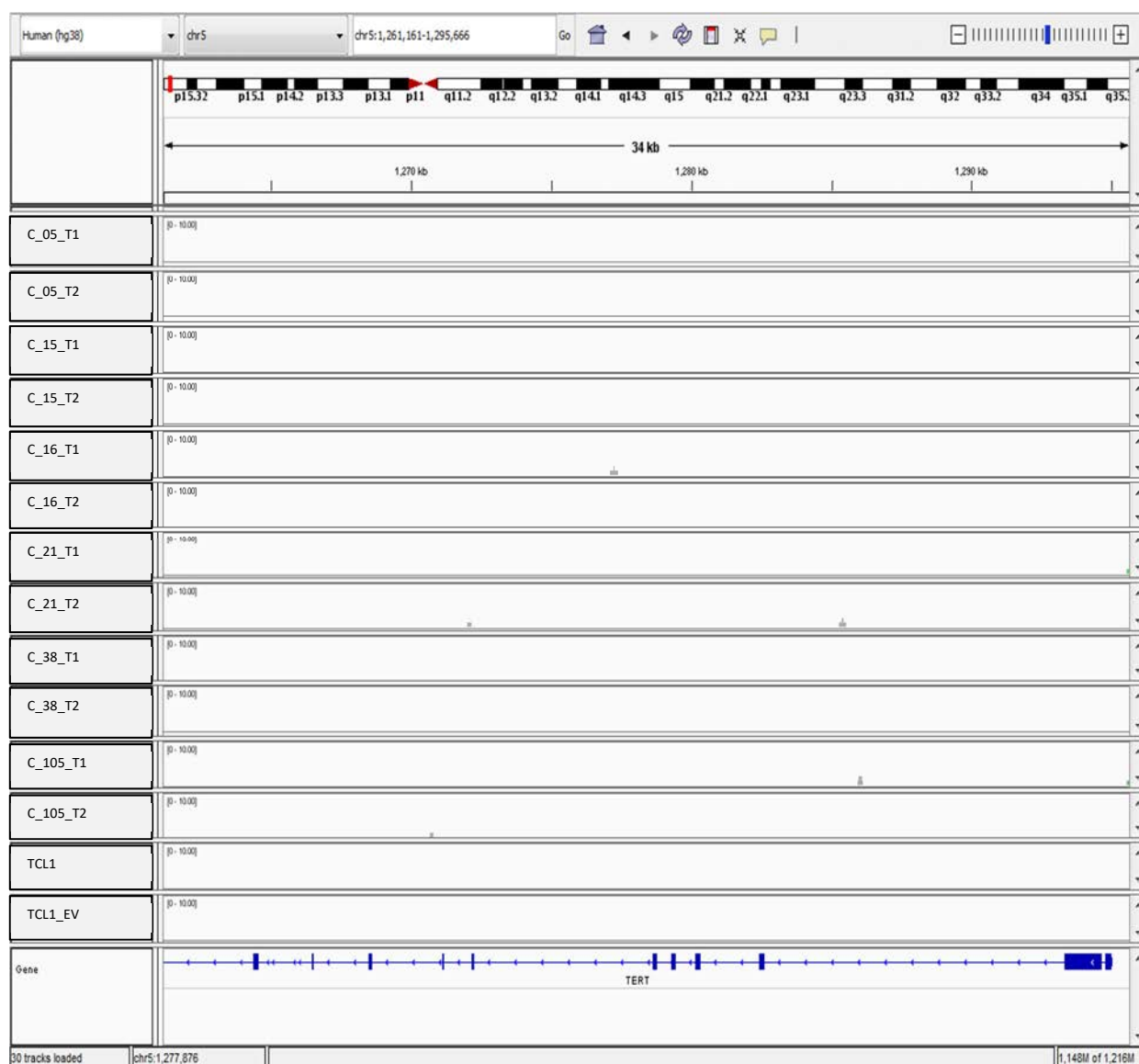
Figure 5.21 No expression of TERT was found in ALT+ post crisis CPs. RNA-seq analysis of the *TERT* gene was visualised in .bam files using IGV. No reads that mapped to TERT exons were identified. Normal SV40-TCL1 and SV40-TCL1-EV were used as controls in this comparison.
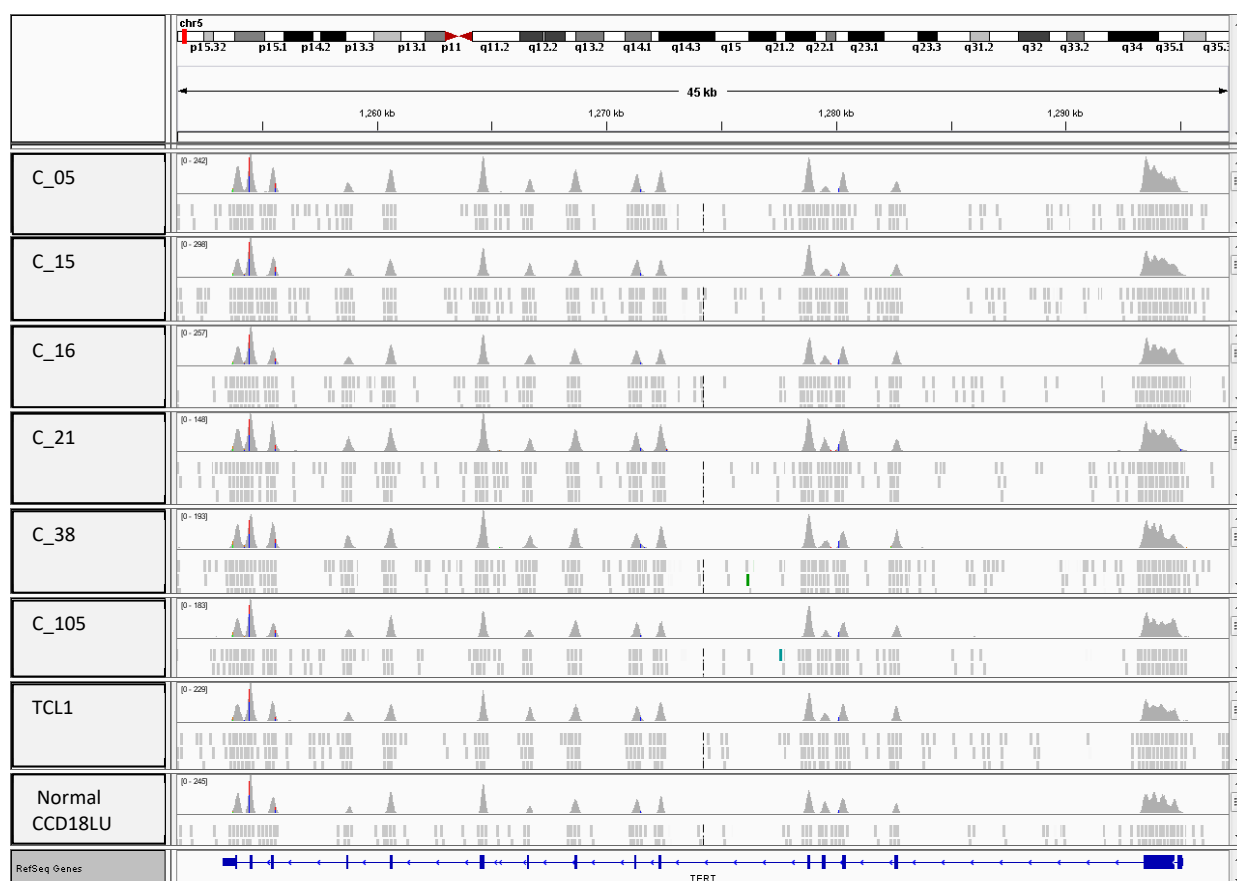
Figure 5.22 Wild type *TERT* sequence was present in all the ALT+ post-crisis CPs. The exonic regions of the *TERT* gene were inspected in WES data generated for each CP, TCL1 and normal CCD18LU cells. The *TERT* annotation and hg19 primary assemble genome r were used as a reference. No somatic SNV in TERT were observed. Normal CCD18LU and SV40-TCl1 were used as controls in this comparison.

5.4 Discussion

5.4.1 Successful ATRX editing in SV40-transformed fibroblast cell line increased the probability of ALT activation

Somatic mutations in *ATRX* or its partner *DAXX* are strongly associated with activation of the ALT mechanism in both cell lines (Lovejoy *et al.*, 2012) and in cancers (Schwartzentruber *et al.*, 2012; Heaphy *et al.*, 2011b). *ATRX* exon 9 is of particular interest as 50% of *ATRX* mutations in gliomas are clustered in this region (Jiao *et al.*, 2012). Therefore, two CRISPR-Cas9 target sequences (CRISPR-TS-1 and CRISPR-TS2) were selected to introduce inactivating mutations in *ATRX* exon 9 in pre-crisis SV40-transformed CCD18LU/TCL1 cells.

In total, 6 of 165 isolated clones gave rise to CPs that emerged from crisis after relatively short time (50-60 days), and all were shown to be ATRX$^{-/-}$. The immortalisation frequency (2.75×10^{-6}) was higher in the CPs, which arose from cells co-transfected with both CRISPR-TS-1 and CRISPR-TS2 than the clonal populations that were transfected with just one target (1.35×10^{-7}). This suggests that CRISPR/Cas9 targeting two site in exon 9 increased the chance of *ATRX* knockout. Although the number of *ATRX* edited clones pre-crisis could not be ascertained (because cells entered the period of crisis before reaching confluency in a 6 well plate), 100% (6/6) of post-crisis CPs showed mutations at the expected sites. Missense prediction analysis showed that all the mutations were expected to trigger nonsense mediated decay (NMD) and ATRX mRNA termination. The absence of ATRX protein was confirmed by IF-staining in early post-crisis cells (T1), and by western blot analysis at late post-crisis (T2).

As there is no direct assay for ALT, several ALT phenotypic markers were analysed. APBs were detected in all post-crisis CPs at the early post-crisis (T1) although the APB frequency and the number of APBs per nucleus varied between the 6 CPs. Similar variation in the abundance of C-circles was also shown in CPs at the late post-crisis (T2). This could indicate variation in the molecular mechanisms that underlie telomere lengthening via ALT. Furthermore, ALT activation in post-crisis CPs was confirmed by telomere length analysis.

The expression of *TERT* has investigated using high depth next-generation sequencing of mRNA. TERT transcripts were not detectable in any post-crisis CP or in the pre-crisis TCL1cells. To rule out inactivation of the *TERT* gene by a mutation, the integrity of the gene has investigated whole exome sequencing data. No somatic or germline inactivating mutation was found in the post-crisis CPs or in the pre-crisis TCL1 cells. These findings were incompatible with data presented by Bower et al., (2012), and Napier et al., (2015). The clonal origin of the SV40-LT transformed TCL1 cells could suggest that the higher frequency of ALT activation in the post-crisis TCL1-ATRX$^{-/-}$ CPs may be due to the accumulation of other genetic or epigenetic mutations in the pre-crisis cells that also increase the chance of ALT activation as cells emerge from crisis. Repeating the same ATRX editing experiment in a different SV40-LT transformed CCD18LU pre-crisis clone might address this unresolved question.

In summary, this work showed that inactivating mutations were successfully introduced into the *ATRX* gene using CRISPR-Cas9 targeting in a pre-crisis clone of SV40-transformed fibroblasts. All the *ATRX$^{-/-}$* post-crisis CPs activated ALT in the presence of wild type *TERT*. As other genetic or transcriptional changes may be required for ALT activation, the genomic instability, genetic and transcriptional alternations at early post-crisis CPs have been characterised in chapter 6.

Chapter 6 Genome-wide analysis of ALT activation in post-crisis CPs

6.1 Introduction

6.1.1 Telomere driven genomic instability, loss of ATRX, and activation of ALT in SV40-transformed cells.

Transfection of primary fibroblast cells with SV40-LT expression plasmid inactivates the p53 and Rb proteins, which allow cells to continue dividing till they enter the crisis (Wright and Shay, 1992; Shay and Wright, 2005). In the absence of a fully functional DNA damage response, cells that containing critically short and unprotected telomeres can undergo fusions creating dicentric chromosomes and initiating breakage-fusion-bridge cycles (Wright and Shay, 1992; Shay *et al.*, 1993; Shay and Wright, 2005). This telomere-fusion driven instability results in various types of genomic rearrangements that are often seen in tumours (Counter *et al.*, 1992; De Lange, 2005; Davoli *et al.*, 2010). Telomere fusion occurs by either the c-NHEJ pathway (c-NHEJ) (Smogorzewska *et al.*, 2002) or the alternate end joining pathway (a-EJ) (Maser *et al.*, 2007). The c-NHEJ pathway was shown to mediate inter-chromosomal telomere fusion events, which result in chromosomal translocations. The a-EJ pathway has been shown to facilitate the intra-chromosomal telomere fusions, which mostly arise between sister-chromatids of short dysfunctional telomeres. In contrast to the c-NHEJ, a-EJ-mediated telomere fusions appear to facilitate escape from crisis (Jones *et al.*, 2014).

The dicentric chromosomes formed from telomere fusions that arise during crisis are associated with mitotic failure, which can also lead to tetraploidization (Davoli *et al.*, 2010; Davoli and de Lange, 2012). Tetraploidization is an intermediate step towards tumorigenesis in many cancers and is followed by chromosome loss (Shackney *et al.*, 1989). In SV40 transformed somatic mouse cells, the persistent DNA damage signalling at short telomeres is associated with prolonged S/G2 and mitotic failure that is followed by re-entry into G1/S, endoreduplication resulting in tetraploidy (Davoli *et al.*, 2010; Davoli and de Lange, 2012). Thus, the consequences of the breakage-fusion-bridge cycles and tetraploidization are genomic instability represented by gains and losses of chromosomes and thereby changes in gene expression (Davoli and de Lange, 2011).

Loss of ATRX expression is also associated with an increased level of genomic instability in cells that use ALT (Lovejoy *et al.*, 2012). This is consistent with the role of ATRX in resolving the G-quadruplex structures at the G-rich sequences, thereby ensuring fidelity during transcription and DNA replication through G-rich regions (Levy *et al.*, 2015; Clynes *et al.*, 2015). In another study, loss of ATRX in Hela TEL+ cells resulted in a prolonged transition from prometaphase to metaphase which was associated with defective sister-chromatid cohesion and clustering of telomeres at the metaphase plate (Ritchie *et al.*, 2008). Micro-nucleation is a mechanism by which cells can lose or re-arrange chromosomes (Crasta *et al.*, 2012). It has been shown that ~68% (15/22) of established ALT+ cell lines exhibited a high level of genomic instability measured by a high frequency of cells with micronuclei (10–30%) in contrast to TEL+ positive cells, such as BJ (SV40-telomerase positive) and HeLa, that showed a low level of genomic instability (<8% of cells with micronuclei) (Lovejoy *et al.*, 2012).

6.2 Aims

In previous chapter I showed that loss of ATRX in SV40-transformed fibroblasts facilitate and increases the probability of ALT activation in cells that survive crisis. In this chapter, integrated analyses of whole-exome sequencing (WES) and transcriptome profiling (RNA-seq) have been conducted to uncover genetic and gene expression changes that are associated with activation of the ALT mechanism in the $ATRX^{-/-}$ crisis surviving clonal populations.

6.3 Results

6.3.1 Gene expression and mutation analyses in ATRX$^{-/-}$ post-crisis CPs

6.3.1.1 Alignment summary of RNA-seq

A total of 18 samples (3 x biological replicates of TCL1, 3 x biological replicates of TCL1_EV, 6 x individual samples at early post-crisis (T1), and 6 x individual samples at late post-crisis T2) were sent to Novogene for non-stranded mRNA sequencing. For each sample there were over 30 million paired-end (PE) reads which passed the Q20 filter, of which 81-85% were uniquely mapped to human genome reference, 1.4-1.8% of the reads mapped to multiple loci and 12.6-16.9% of the reads did not map to the human genome. The uniquely mapped reads were equally distributed to forward and reverse strands in all the samples (Appendix 6.1). The comparative transcription analysis was conducted by counting 83-87% of the reads that mapped uniquely to genes.

6.3.1.2 Cluster analysis of gene expression profiles in ATRX-/- post-crisis CPs.

Similar to RNA-seq analysis in chapter 4, a gene counts matrix was generated of uniquely mapped PE reads for gene quantification by FPKM normalisation and deferential expression analysis using DESeq2. To track the changes in gene expression between early and late post-crisis, samples were clusters by their expression profiles using correlation analysis. The dendrograms were then generated by hierarchical clustering in complete linkage. The dendrograms showed that the pre-crisis TCL1 and TCL1_EV replicates cluster together, while post-crisis clonal populations clustered into two distinguishable subgroups post-crisis A, post-crisis B as shown in Figure 6.1. At post-crisis T1, the subgroup A included CP_05, CP_15, CP_16, while the subgroup B included CP_21, CP_38, and CP_105. At post-crisis T2, the expression profile in CP_16 diverged from others in subgroup A but showed a greater similarity to the expression profiles of sample in subgroup B, whereas the expression profiles of CP_15 and CP_05 retained their similarity at T2 in the post-crisis A group (Figure 6.1). This could suggest that the various subclones that contributed to the CP_15 and CP16 as they emerged from crisis changed between T1 and T2. For example subclones with favourable traits, required for survival, and may have outgrown other subclones within CP_15 and CP_16 between the time points.
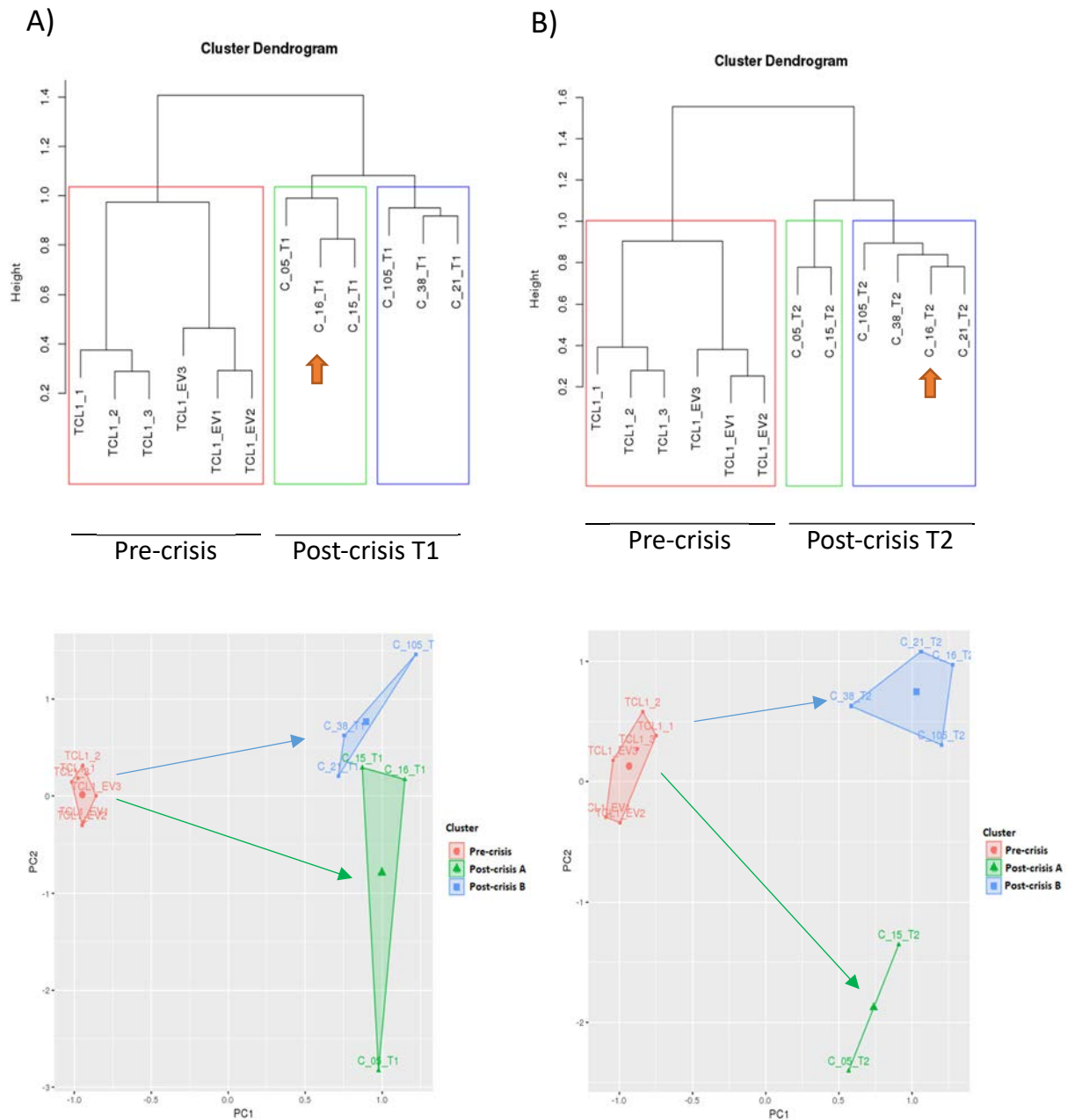
Figure 6.1 Sample-sample correlation analysis based on gene expression profiles. The correlation distance measurements between samples were calculated from log-cpm of all expressed genes. Cluster dendrograms were generated by hierarchical clustering in complete linkage, (A) dendrogram and PCA analysis of pre-crisis TCL1, TCL1-EV replicates and post-crisis clonal populations at T1, (B) dendrogram and PCA analysis of the same pre-crisis TCL1, TCL1-EV replicates with the post-crisis clonal populations at T2.

6.3.1.3 Phenotypes associated with the changes in gene expression in *ATRX*-/- post-crisis CPs

To determine whether the changes in the expression profiles between early and late post-crisis clonal populations were associated with phenotypic markers of genome instability, micronuclei analysis was performed as described in section 2.2.2.2.4. At the early post-crisis timepoint (T1), cells in CP_05, CP_15, and CP_16 showed relativly high levels of genomic instablity (23- 26% of cells with micronuclei), while cells in CP_21, CP_38, and CP_105 showed a lower level of genome instablity (15-16% cells with micronuclei). At the later post-crisis timepont (T2), the frequency of micronuclei had reduced to 12-13% in CP_05 and CP_15; 10% in CP_16; 5-7% in CP_21, CP_105; and a smaller reduction to ~12% in CP_38 (Figure 6.2 and Table 6.1). The lack of a substantial reduction in micronuclei frequency in CP_38 might be explained by clonal homogeneity, as the cells in this population emerged from just a single colony that survived crisis (see Table 5.3 in chapter 5). In summary four of the six late post-crisis clonal populations continued to show slightly increased frequencies of micronuclei (10-13%) (Figure 6.2).

The micronuclei analysis was consistent with the gene expression clustering at the T1, as CP_05, CP_15, CP_16 showed similar expression profiles and had higher levels of genomic instability (HGI), whereas CP_21, CP_38, CP_105 showed lower genomic instability (LGI) and shared similar expression profiles.

Table 6.1 Summary of micronuclei analysis in ATRX-/- post-crisis CPs. The frequency of cells with micronuclei was obtained by manual inspection and counting of interphase DAPI-stained nuclei at early post-crisis (T1) and late post-crisis(T2).

ID	Time point	Number of coverslips screened	Total number of screened nuclei	% cells with micronuclei
TCL1	Pre-crisis	2	237	8
CP_05	T1	2	201	23
	T2	3	336	12
CP_15	T1	3	347	25
	T2	3	410	13
CP_16	T1	2	201	26
	T2	3	397	10
CP_21	T1	2	219	16
	T2	3	384	5
CP_38	T1	2	311	15
	T2	3	341	12
CP_105	T1	2	311	15
	T2	3	374	7
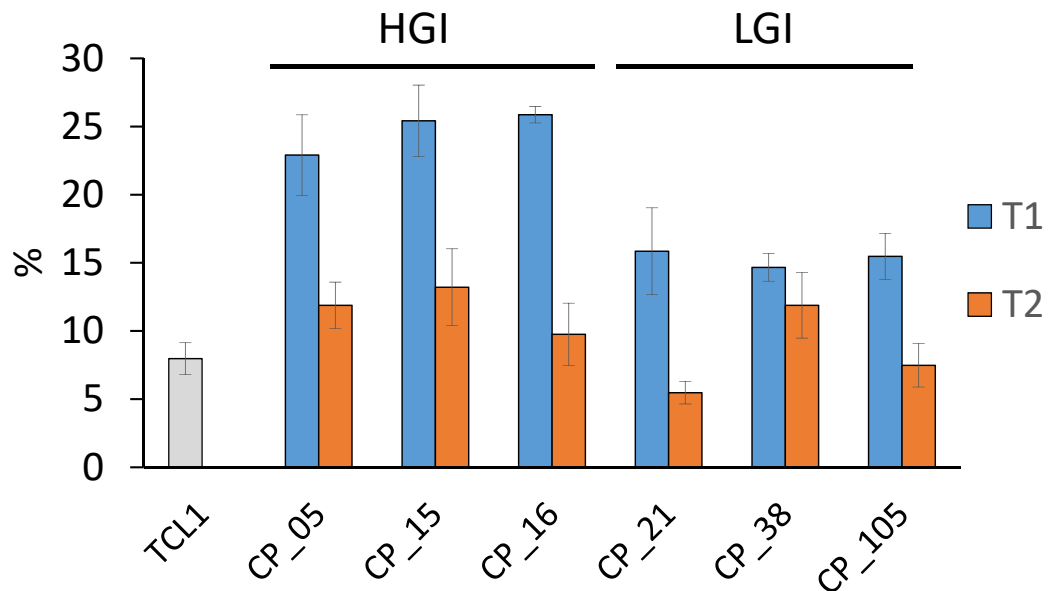
Figure 6.2 Frequency of micronuclei in ATRX-/- post-crisis CPs. HGI= post-crisis clonal populations with relatively high genomic instability at T1. LGI=post-crisis clonal populations with relatively low genomic instability at T1.

6.3.1.4 Summary of the differential gene expression analysis between pre-crisis groups and post-crisis subgroups

To identify the changes in gene expression that may be associated with activation of the ALT mechanism, the differential expression analysis was conducted in multiple comparisons using DESeq2 tool. The analyses included: comparison of gene expression profiles between the pre-crisis groups (TCL1_EV and TCL1); comparison of each of the four post-crisis subgroups (HGI_T1, HGI_T2, LGI_T1, and LGI_T2) with the pre-crisis (TCL1) group. To understand the changes in the expression profiles better, the number of DE genes was calculated using different thresholds of log2 fold changes (log2 fold change >0, >1, >2, >3). The data presented in Table 6.2 showed that only a few genes (total 95 genes) showed DE between the pre-crisis TCL1_EV and TCL1 samples. This similarity in expression profiles was expected as both sample of cells are from pre-crisis cells, although as a consequence of transfection and cloning with the empty vector, the cells in the TCL1_EV samples were~10 PDs closer to the crisis than the TCL1 samples. In comparison with pre-crisis TCL1, the number of DE genes are 2319 and 2562 in LGI_T1 and HGI_T1, rising to 3361 and 3703 in HGI_T2 and LGI_T2. The increase in the number of DE genes at late post-crisis in comparison to early post-crisis suggests more changes occurred during the expansion of the CPs.

Table 6.2 Summary of DE analysis between post-crisis subgroups and TCL1. It showed the number DE genes at different cut-off points of log2 fold change and FDR <0.05

Comparison	Status	Number of DE genes, log2 fold change			
		>0	>1	>2	>3
TCL1_EV vs TCL1	Down	69	53	26	14
	Up	26	21	9	6
	Total	95	74	35	20
HGI_T1 vs TCL1	Down	1302	802	280	96
	Up	1260	581	338	213
	Total	2562	1383	618	309
HGI_T2 vs TCL1	Down	1622	1076	479	222
	Up	1739	866	318	183
	Total	3361	1942	797	405
LGI_T1 vs TCL1	Down	1183	745	261	90
	Up	1136	512	272	155
	Total	2319	1257	533	245

LGI_T2 vs TCL1	Down	1772	1011	372	136
	Up	1931	1009	557	314
	Total	**3703**	**2020**	**929**	**450**

The DE genes were also shown in volcano plots (Figure 6.3). The top 20 highly differently expressed genes are labelled in the volcano plots (based on p value). The downregulated genes in both HGI_T1 and LGI_T1 include two members of Myxovirus (Influenza) resistance homolog proteins (MX1/2) , two members of Interferon Alpha Inducible protein (IFI6, IFI27), a member of Interferon Induced Transmembrane (IFITM1), Signal Transducer and Activator of Transcription 1 (STAT1), Poly(ADP-Ribose) Polymerase Family proteins (PARP9), genes involved in the fatty acid and protein metabolism such as Stearoyl-CoA Desaturase(SCD), and Proprotein Convertase 9 (PCSK9), sugar metabolism Carboxypeptidase Q (CPQ), cell membrane associated genes such as Lymphocyte Antigen 96 (LY96), and Phospholipid Scramblase 1 (PLSCR1), as well as GTP-binding protein (GEM), and Polypeptide N-Acetylgalactosaminyltransferase 12 (GALNT12). On the other hand, the labelled upregulated genes in both HGI_T1 and LGI_T1 are RAP1GAP2 (RAP1 GTPase Activating Protein 2), Butyrylcholinesterase (BCHE), and Solute Carrier Family 16 Member 14 (SLC16A14). Most of down and up regulated genes labelled in T1 are also DE in T2, but with some changes. For example, CPQ and SCD were observed in the list of top 20 DE genes in T1, but not in T2. Whereas the opposite was observed in the case of the NR4A3 and STE20-Related kinase Adaptor protein Beta (STRADB), as they were not found in the list of top 20 in T1 but present in T2.
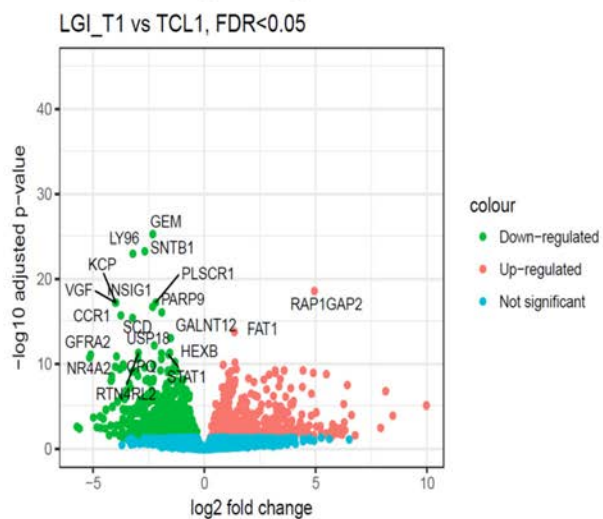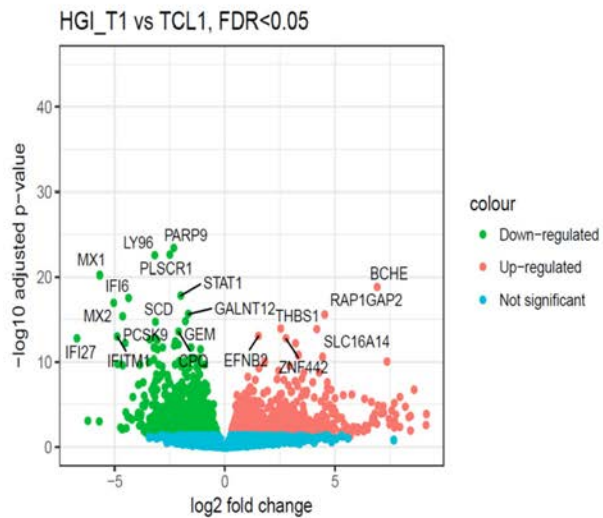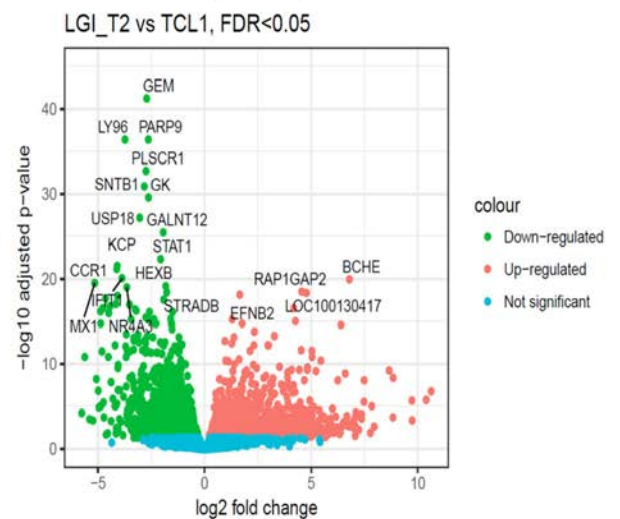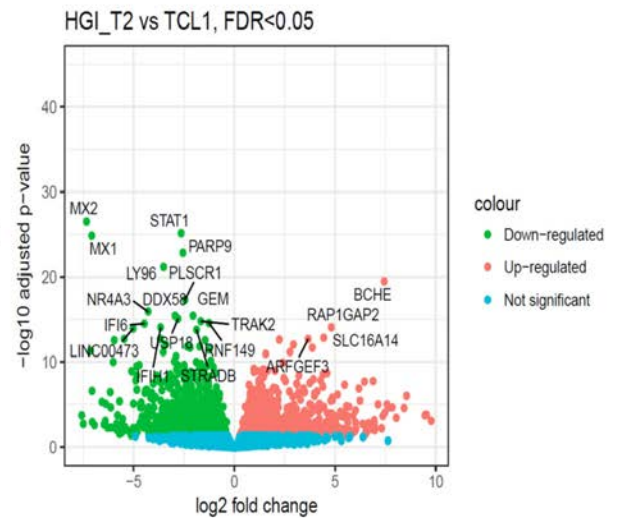
Figure 6.3 Volcano plots of DE genes from the comparison of post-crisis subgroups against TCL1.The top 20 highly differently expressed genes (based on p-value) are labelled. The x-axis showlog2 fold change in gene expression, and the y-axis show –log10 of adjusted p-value.

6.3.1.5 Gene ontology analysis of DE genes between post-crisis and pre-crisis groups

To characterise the biological process that are altered as a results of changes in gene expression in the early post-crisis time point, GO enrichment analysis was preformed using DE genes identified at early post-crisis from the comparison of HGI_T1 vs TCL1 (Appendix 6.2) and LGI_T1 vs TCL1 (Appendix 6.3). The core of enrichment genes in biological processes in both HGI_T1vsTCL1 (Figure 6.4) and LGI_T1vs TCL1 (Figure 6.5) showed downregulation of genes that are related to cell innate immunity response to viruses including genes that can be induced by response to type I /alpha interferons (IFNs), as well as blood vessel morphogenesis (angiogenesis) and proliferation. The DE genes related to interferon response are members of Interferon Induced Transmembrane Proteins (IFITMs), members of Interferon Induced Protein with Tetratricopeptide (IFITs), two members of Myxovirus (Influenza) Resistance homolog proteins (MX1/2), members of Interferon Alpha Inducible Proteins (IFIs), members of Interferon Regulatory Factor 7 (IRF9/7), and member of the Tripartite Motif family proteins (TRIMs), and some interferon-stimulated genes (such as ISG20), and SP100. Furthermore, this pathway also included an interferon induced RNase, 2'-5'-Oligoadenylate Synthetase 1 (OAS1), and Signal Transducer and Activator of Transcription 1 (STAT1). This suggests that the JAK/STAT signalling pathway is down regulated in both subgroups of early post-crisis CPs.

Type I interferons are a subclass of cytokines present in most cells and their expression levels are elevated in response to viruses. Binding the IFNs to their specific cytokine receptors activates signal transduction by the JAK-STAT signalling pathway (Aaronson and Horvath, 2002). The activated JAKs phosphorylate tyrosine residues on their receptors, creating binding sites for STATs proteins through the SH2 domains (Jatiani *et al.*, 2010). The activated STATs interact with Interferon Regulatory Factor 9 (IRF9) to form a complex called ISGF3 that moves to the nucleus. Inside the nucleus, the ISGF3 binds to specific nucleotide sequences termed IFN-stimulated response elements (ISREs) in the promoters of ISGs genes. This can result in an antiviral defence status, which may also activate apoptosis signalling (Barber, 2000). The ISGs include large set of genes, many of them are localised in PML nuclear bodies. *SP100*, and *ISG20* are the best characterised genes that contain IFN-stimulated response elements in their promoter

regions, making them primary targets for the ISGF3 complex (Stadler *et al.*, 1995; Grotzinger *et al.*, 1996; Gongora *et al.*, 1997). Furthermore, it has been shown that expression of STAT1 is enhanced by PARP9 (Zhang *et al.*, 2015). Reduction in expression of STAT1 seems to be associated with the downregulation of type 1 interferons that also results in reduced expression of a large group of genes that explains the downregulation in *SP100*, *ISG20*, *MX1/2*, and other genes. Interestingly, the core of enriched genes also showed reduction in the actin alpha-Actin-2 (*ACTA2*) in HGI_T1, but not LGI_T1.

However, the other gene expression changes (up and down) in both post-crisis subgroups were in genes involved in blood vessel morphogenesis. These include downregulation of two transcription factors of NR4A family genes (*NR4A1*, *NR4A3* in HGI; and *NR4A1, NR4A2, NR4A3* in LGI). The NR4A family genes play roles in metabolism and cancer pathways such as angiogenesis, proliferation, apoptosis, and DNA repair (Mohan *et al.*, 2012; Ranhotra, 2015). For example, NR4A2 has been shown to mediate VEGF-induced angiogenesis, and increase the expression of proliferation markers (Zhao *et al.*, 2011). NR4A2 showed anti-apoptotic function by inhibiting p53-mediated apoptosis signalling (Riggins *et al.*, 2010). NR4A2 has also been shown to localise to DSB sites after UV exposure which predicts a potential role in a DNA repair pathway (Yin *et al.*, 2017). In mouse tumour models, NR4A1 has been implicated in reducing cell proliferation by inhibition of Wnt/B-catenin signalling (Chen *et al.*, 2012). Furthermore, the NR4A1 protein can initiate the apoptotic cascade via direct association with the apoptotic regulator Bcl-2 in the mitochondria (Thompson and Winoto, 2008). In pancreatic tumours, NR4A1 has also shown a pro-apoptotic effect by inducing the expression of pro-apoptotic genes (Yoon *et al.*, 2011). In another example, NR4A1 interacts with the DNA-PK catalytic subunit and facilitates DSB break repair by classical NHEJ. In extraskeletal myxoid chondrosarcomas, chromosomal fusion between Ewing sarcoma region-1 (EWS) and NR4A3 (chr9:chr22) has been implicated in the upregulation of PPARG which promotes oncogenesis (Filion *et al.*, 2009). Thus, the reduced expression of NR4A1 in crisis surviving CPs appears to avoid the pro-apoptotic effect of this nuclear receptor, but it may also inhibit NHEJ. On the other hand, the upregulation of the members of GATA binding proteins (*GATA3*, *GATA6*) that work as transcriptional activators for the Insulin Like Growth Factor 1 (IGF-1) receptor signalling

and the proto-oncogene gene C-MYC network may be required to prompt the growth and survival signalling (van Hamburg *et al.*, 2008).

To track the changes in gene expression from T1 to T2 in the post-crisis subgroups, two comparative GO enrichment analyses were performed using the clusterprofiler tool. The comparative GO enrichment analysis (Figure 6.6 and Appendix 6.4) showed that the early events of post-crisis (represent by high enrichment score at T1) are downregulation of genes involved in G-protein coupled receptor binding (such as *CCL*s and *CXCL*s), glycoprotein binding (Such as *CTSK, CTSS, FBXO6*) and cytokine binding (such as Such as *IL23R, CSF3R, IL6R*). Whereas the late events included upregulation of transcriptional activators such NR1D1, members of the SOX gene family (*SOX12, SOX9* in HGI_T2 group; and *SOX4* in LGI_T2 group), a member of Forkhead Box gene family (*FOXC2*), Early growth response genes (*EGR1* in HGI_T2 group), Lymphoid enhancer-binding factor 1 (*LEF1*), as well as modelling the TGF-Beta signalling by upregulation of two members of SMAD family genes (*SMAD6* and *SMAD7*).

Although downregulation of gene expression in G-protein coupled receptor binding and glycoprotein binding genes seemed to occur at the early post-crisis time point (T1), it did not seem to persist to T2. These genes coded for transmembrane receptors that serve as ligand binding for several chemical compounds, odors, pheromones, and chemokines that can mediate JAK/STAT activation in some lymphocyte cells (Wong and Fish, 1998).
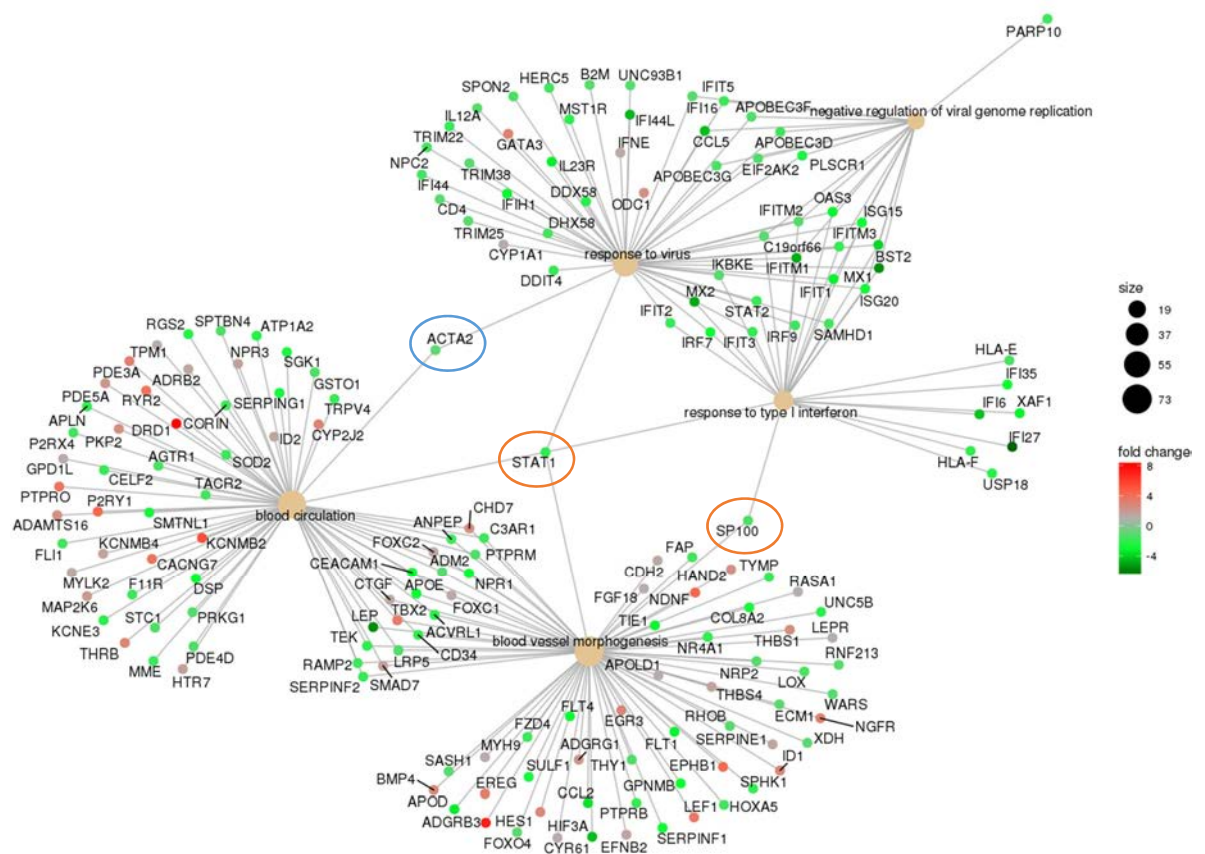
Figure 6.4 Visualisation of DE genes in most significantly enriched biological processes associated with HGI_T1 vs TCL1. It illustrates the biological complexities in which a gene (such as *SP100, STAT1*, and *ACTA2*) may belong to multiple biological terms. It also shows the terms linked to the most significant biological processes associated with genes that show DE at HGI_T1. Fold change >2, FDR <0.05.
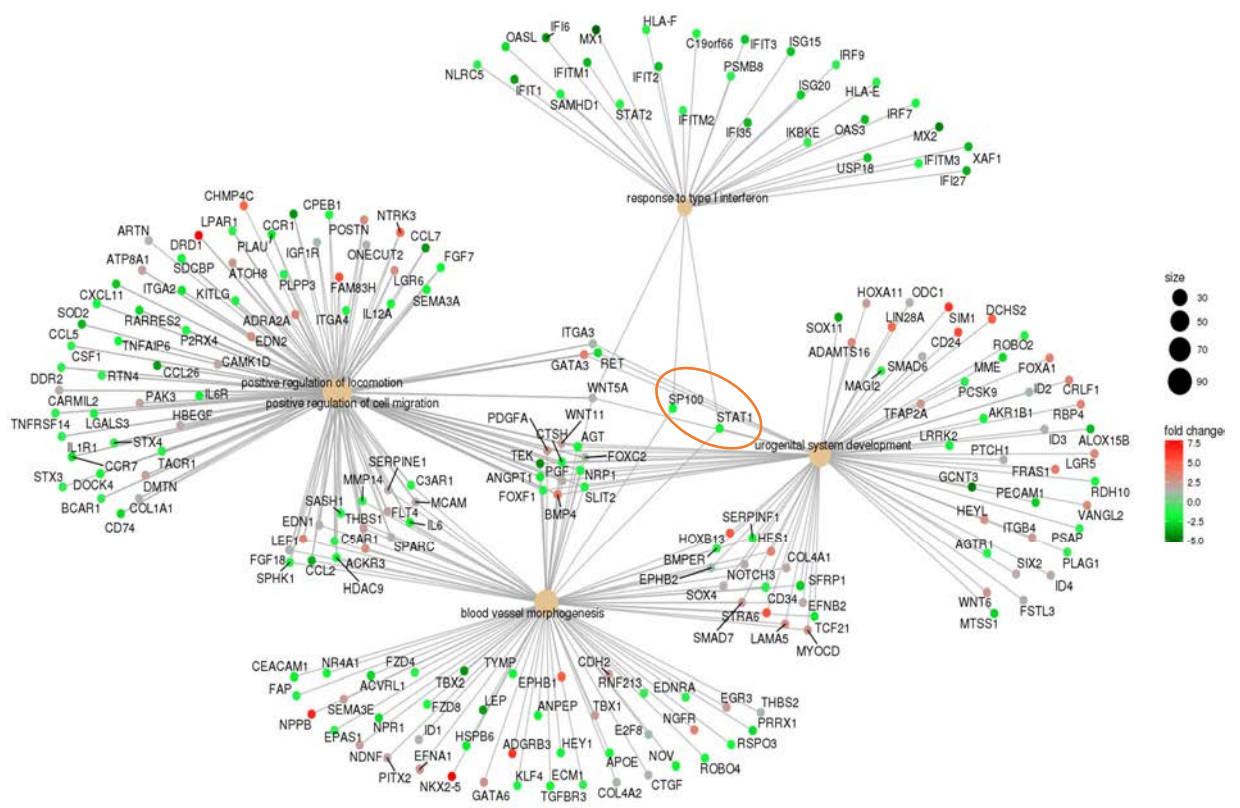
Figure 6.5 Visualisation of DE genes in most significantly enriched biological processes associated with LGI_T1 vs TCL1. It illustrates the biological complexities in which a gene may belong to multiple biological processes, it also shows the most significant terms associated with the biological processes altered in LGI_T1. Fold change >2, FDR <0.05.
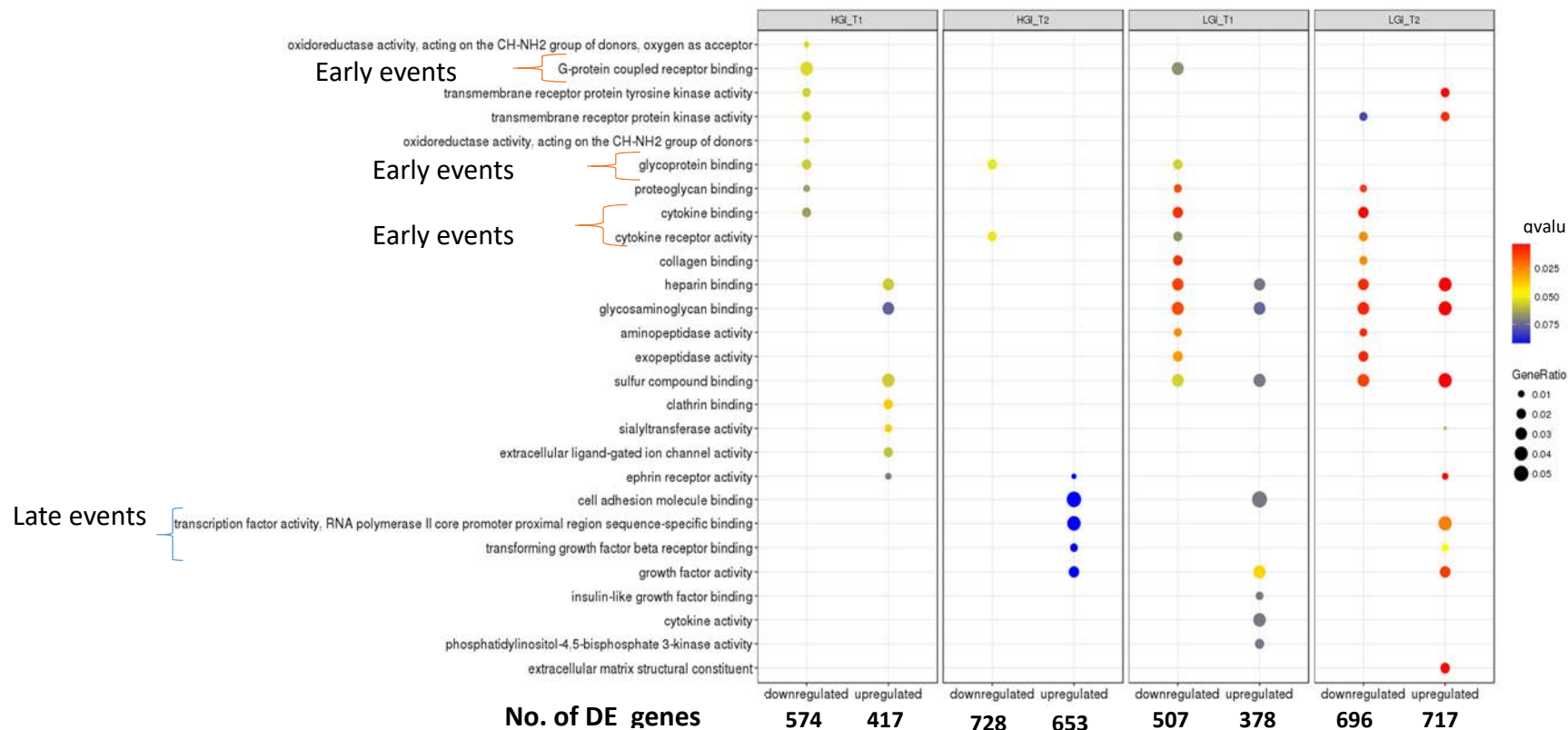
Figure 6.6 Comparative GO enrichment analysis of DE genes in subgroups at T1 and T2. It illustrates comparison of GO terms in four paired comparisons: HGI_T1vsTCL1, HGI_T2 vsTCL1, LGI_T1 vsTCL1, LGI_T2 vsTCL1. Log2 fold change >1, FDR <0.05. The dot size represents the ratio of gene enrichment, while the colour gradient (from blue to red) was used to show the q value of enrichment comparison between different subgroups.

To predict the functions of unknown genes, cluster analysis of genes with similar expression patterns was carried out by the hierarchical clustering of the top 1000 highly variable genes (Figure 6.7). The hierarchical clustering resulted in ten subclusters with 28 to 210 genes for each subcluster. In comparison with pre-crisis TCL1 and TCL1_EV, subcluster 1 showed upregulation in LGI but not in HGI, while subclsuters 2-4 showed downregulation in HGI_T1 and more in HGI_T2, but not in in LGI subgroups. Subcluster 5 showed upregulation in both HGI and LGI (Figure 6.8). Subcluster 6 and 7 showed upregulation in all post-crisis subgroups, except LGI_T1. Furthermore, subcluster 8 showed downregulation in HGI and LGI at both T1 and T2 (Figure 6.8), while subclusters 9 and 10 showed more variability between HGI and LGI subgroups.

At log2 fold change > 4, 22 genes with unknown function were included. Six of 22 are found in both HGI_T1 and LGI_T1, which suggests that they play similar roles in the HGI and LGI subgroups. Five genes (ENSG00000206532, ENSG00000227706, ENSG00000275216, ENSG00000276850, ENSG00000279720) are found in the subcluster 5 and they are predicted to play role in GABA receptor activity, while one gene (ENSG00000228318) in the subcluster 8 is predicted to encode for specific subtype of cytokine receptors that can activate the TNF signalling pathway (Table 6.3). However, the function of three genes (ENSG00000231083 /subcluster 1, ENSG00000224932/ subcluster 7, and ENSG00000260918/subcluster 10) cannot be predicted as a results of no enrichment for the genes in the subclusters.
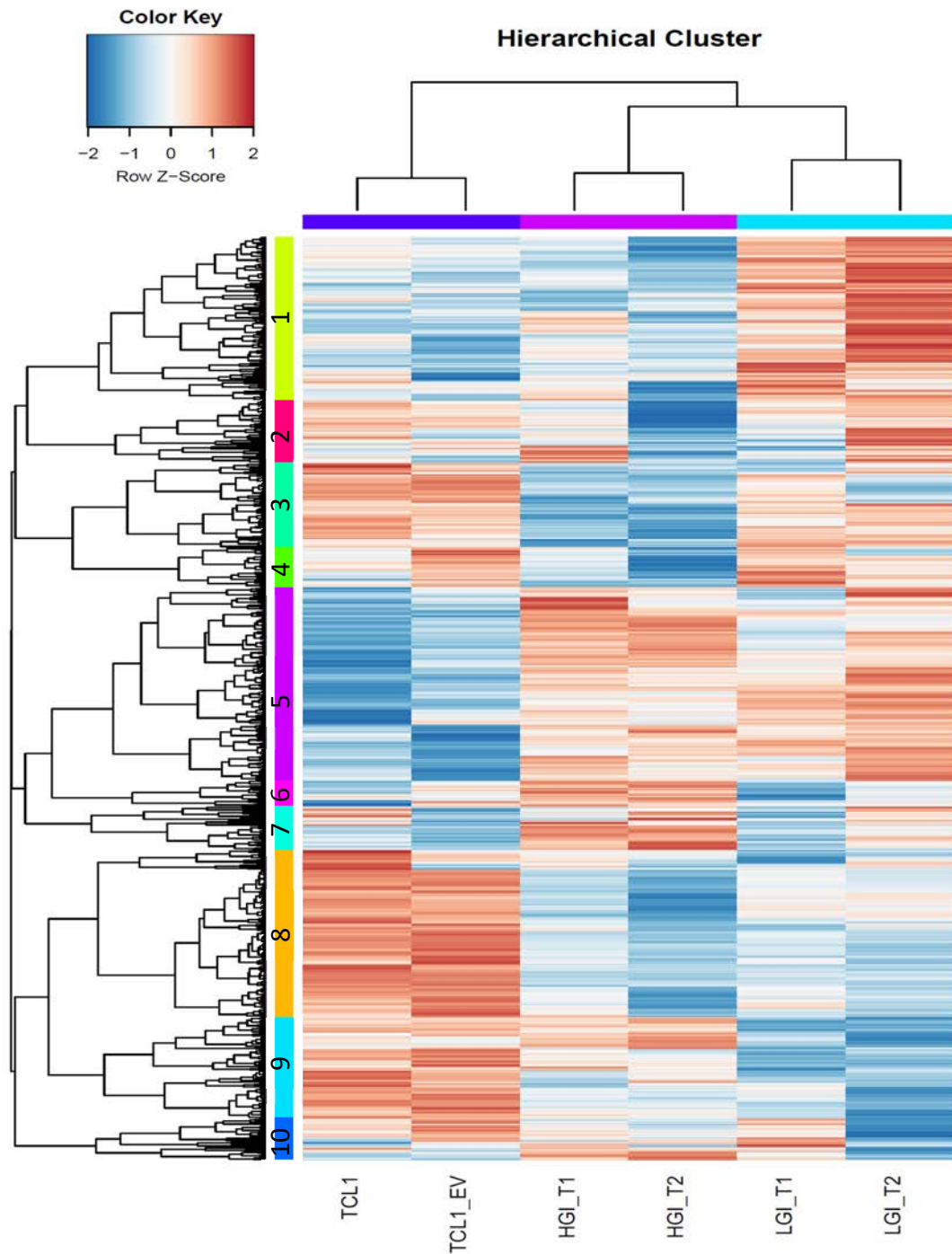
Figure 6.7 Heatmap of the genes that showed expression changes between groups of samples.The cluster analysis of the genes with similar expression patterns was carried by applying the hierarchical clustering with complete linkage on log2(FPKM+1). Ten subclusters were highlighted in different colours on the left sidebar of the heatmap.
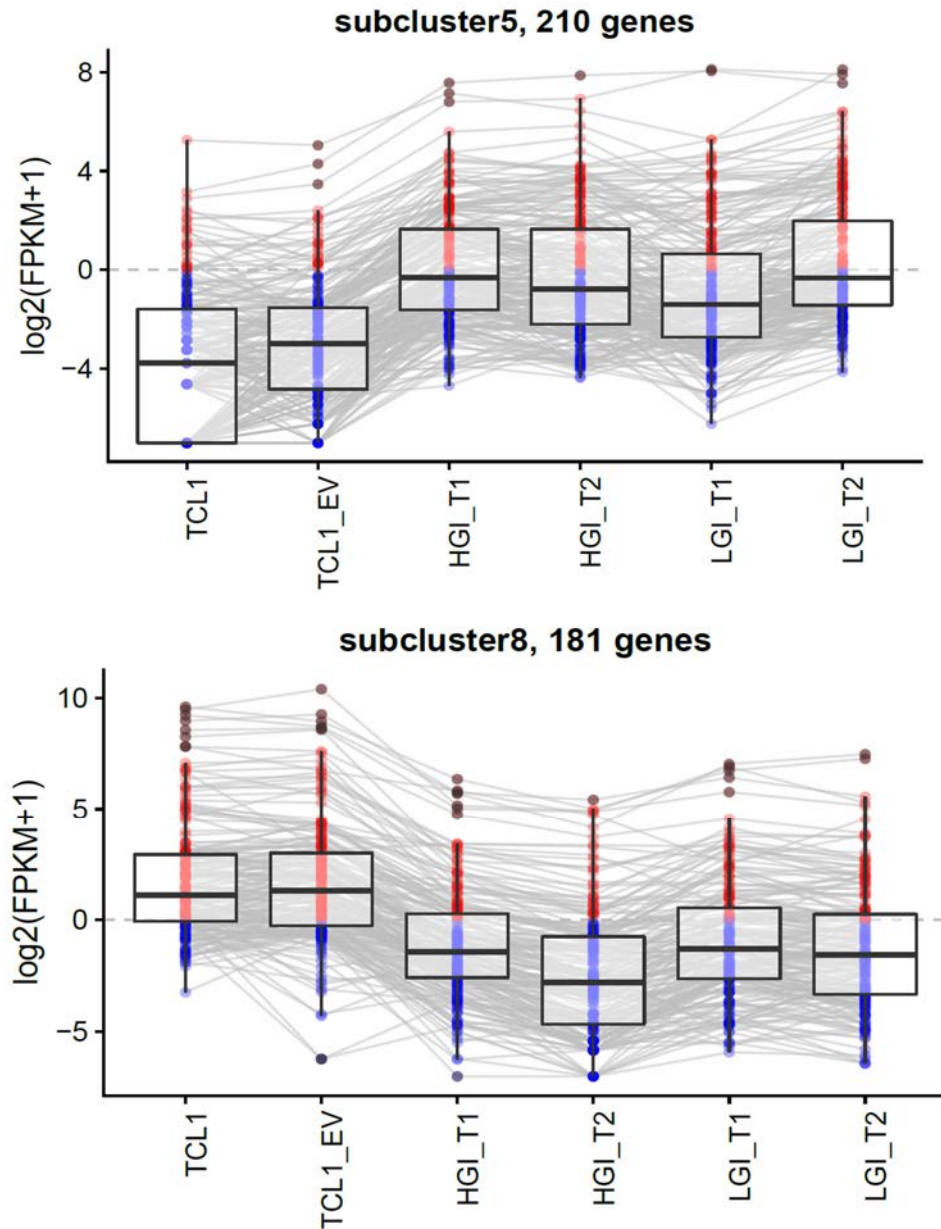
Figure 6.8 Visualising the expression of the genes in two selected subclusters that were identified by hierarchical clustering. Each grey line represents the relative expression value (log2(FPKM+1).The x-axis shows sample groups and the y-axis corresponds to the relative expression value. Red dots represent genes with relative expression > 0, while the blue dots represent genes with the relative expression <0. The distribution of relative expression values was overlaid by box blots that show the median, upper and lower intervals for overall relative expression in each sample group.

Table 6.3 Gene function prediction analysis of uncharacterised genes that showed highly variable expression in post-crisis subgroups at T1. Log2 fold change > 4. Ensemble Gene ID is highlighted in yellow for genes shown in both HGI_T1vs TCL1 and LGI_T1vs TCL1.

Ensambl Gene ID	Comparison	log2Fold Change	p value	p adj	Sub-cluster	Subcluster size (# of genes)	GO enrichment	KEGG enrichment
ENSG00000231083	LGI_T1vsTCL1	4.174795	2.79E-05	8.82E-04	1	177	No sig enrichment	No sig enrichment
ENSG00000206532	HGI_T1vsTCL1	5.959756	4.28E-07	2.63E-05	5	210	GABA receptor activity Endopeptidase inhibitor activity	No sig enrichment
ENSG00000206532	LGI_T1vsTCL1	4.037423	6.35E-04	9.47E-03	5			
ENSG00000211772	HGI_T1vsTCL1	5.510485	3.79E-03	3.20E-02	5			
ENSG00000227706	HGI_T1vsTCL1	6.765467	2.08E-12	5.75E-10	5			
ENSG00000227706	LGI_T1vsTCL1	7.097157	1.60E-13	1.03E-10	5			
ENSG00000272872	HGI_T1vsTCL1	4.64753	1.20E-04	2.37E-03	5			
ENSG00000273301	LGI_T1vsTCL1	5.522866	1.53E-03	1.78E-02	5			
ENSG00000275216	HGI_T1vsTCL1	6.038103	4.82E-06	1.95E-04	5			
ENSG00000275216	LGI_T1vsTCL1	5.544131	2.69E-05	8.56E-04	5			
ENSG00000276850	HGI_T1vsTCL1	4.548427	6.79E-04	9.05E-03	5			
ENSG00000276850	LGI_T1vsTCL1	4.333451	1.23E-03	1.51E-02	5			
ENSG00000279633	HGI_T1vsTCL1	4.092035	1.93E-03	1.95E-02	5			
ENSG00000279720	HGI_T1vsTCL1	7.030274	4.09E-06	1.69E-04	5			
ENSG00000279720	LGI_T1vsTCL1	6.177366	5.63E-05	1.54E-03	5			
ENSG00000250657	HGI_T1vsTCL1	9.311356	3.89E-05	1.01E-03	6	28	Glutamate receptor activity	Neuroactive ligand-receptor interaction
ENSG00000224932	HGI_T1vsTCL1	5.779908	9.66E-04	1.17E-02	7	47	No sig enrichment	No sig enrichment

ENSG00000203446	LGI_T1vsTCL1	-4.45191	2.01E-16	2.59E-13	8	181	**Cytokine activity Cytokine receptor binding Chemokine receptor binding TNF signalling pathway**	**Cytokine-cytokine receptor interaction Rheumatoid Arthritis IL-17-signalling pathway TNF signalling pathway Chemokine signalling pathway**
ENSG00000225886	HGI_T1vsTCL1	-5.28481	3.93E-14	1.91E-11	8			
ENSG00000228318	HGI_T1vsTCL1	-6.86348	5.65E-06	2.18E-04	8			
ENSG00000228318	LGI_T1vsTCL1	-5.01912	2.99E-04	5.51E-03	8			
ENSG00000254403	LGI_T1vsTCL1	-4.72352	3.69E-06	1.75E-04	8			
ENSG00000254743	LGI_T1vsTCL1	-4.41666	5.57E-05	1.53E-03	8			
ENSG00000270160	HGI_T1vsTCL1	-5.26431	1.52E-05	4.82E-04	8			
ENSG00000203434	LGI_T1vsTCL1	-4.7704	1.28E-04	2.80E-03	9	110	**Cytokine activity Cytokine receptor binding Glycosaminoglycan binding Herparin binding**	**Cytokine-cytokine receptor interaction Influenza A**
ENSG00000224397	LGI_T1vsTCL1	-5.44034	3.13E-05	9.58E-04	9			
ENSG00000232679	LGI_T1vsTCL1	-4.81966	6.16E-13	3.17E-10	9			
ENSG00000255465	LGI_T1vsTCL1	-4.48831	1.44E-07	1.15E-05	9			
ENSG00000260918	HGI_T1vsTCL1	4.373124	3.51E-08	3.32E-06	10	45	No sig enrichment	No sig enrichment

The relationship between DE genes in HGI_T1 and LGI_T1 and the NHEJ pathways (see Appendix 6.5) was investigated by quadruple intersection (Figure 6.9 A and Table 6.4). The results showed that two genes required for c-NHEJ were highly downregulated (log2 fold change >1) in both post-crisis CPs subgroups. These genes are *NHEJ1/ XLF* and *NR4A1*. NR4A1 interacts with the DNA-PK catalytic subunit and facilitate the DSB break repair by c-NHEJ (Malewicz *et al.*, 2011). NHEJ1/ XLF has been shown to interact with XRCC4-LIG4 (Ahnesorg *et al.*, 2006), Ku proteins (Yano *et al.*, 2008), and DNA-PK (Andres *et al.*, 2007). Several studies have shown that NHEJ1/ XLF is required for the ligation step in C-NHEJ pathway (Callebaut *et al.*, 2006; Bhargava *et al.*, 2018). On other hand, *LIG3* was upregulated in HGI_T1, while *POLQ* and *XRCC1* both were upregulated in LGI_T1.

DE and ALT related genes were investigated in HGI_T1 and LGI_T1 by triple intersection (Figure 6.9 B and Table 6.5). It showed that *SP100* was the most significantly downregulated among the ALT related genes (log2 fold change > 1) in both post-crisis CPs subgroups (Figure 6.10). It was shown before that ectopic expression of SP100 sequestered the MRN complex away from PML bodies (W. Q. Jiang *et al.*, 2005). Interestingly, there was a reduction in *MRE11* and the *RAD51D* in both HGI_T1 and LGI_T1 that was associated with upregulation of RAD17 (Table 6.5). *RAD17* is the clamp loader for the 9-1-1 (RAD9-RAD1-HUS1) complex that has been shown to associate with dysfunctional telomeres (Nabetani *et al.*, 2004). This was associated with upregulation of *HUS1*, and *SLX4* in HGI_T1, while an increase in expression of *H2AFX, BRCA2, RPA1*, and *MND1* genes were shown in LGI_T1. This suggests two different DNA repaired mechanisms associated with activation of ALT at the early post-crisis stage. However, upregulation of *BLM, FANCD2, and FANCM* helicases (Table 6.5) may be associated with the increase of homologous recombination activity at telomeres. Furthermore, it was shown that BLM promotes long tract telomere extension by participating in dissolution of Holliday junctions with non-crossover events (Sobinoff *et al.*, 2017).

A)
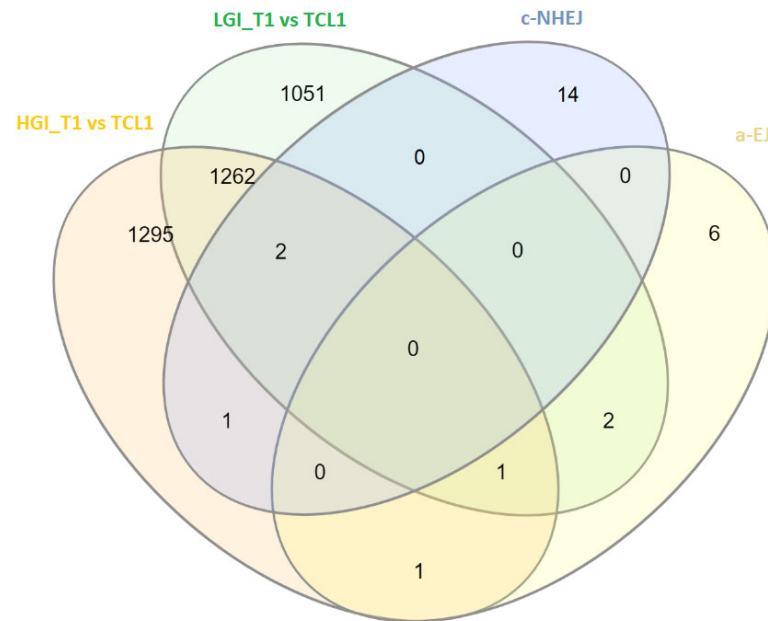
B)
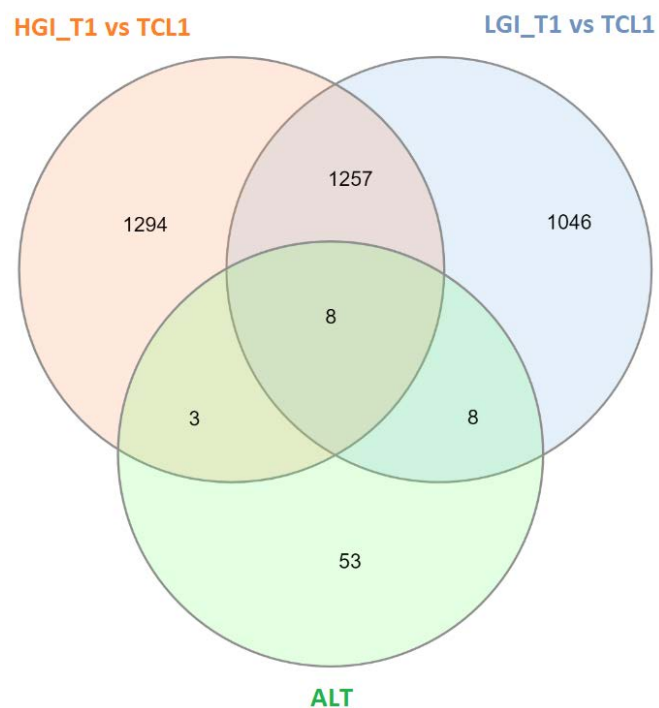
Figure 6.9 Intersection of DE genes with NHEJ and ALT related genes. (A) Intersection of DE genes identified from HGI_T1 vs TCL1and LGI_T1 vs TCL1 with c-NHEJ and a-EJ related genes, (B) Intersection of DE genes from HGI_T1 vs TCL1 and LGI_T1 vs TCL1 with ALT related genes. All significantly DE genes (FDR<0.05) were included in this analysis.

Table 6.4 DE genes associated with NHEJ pathways. Genes with absolute log2 fold change > 1 shaded in yellow. FDR<0.05.

c-NHEJ pathway				
Gene Symbol	ENSMBL Gene ID	Sig-DE in	Log2 FC HGI_T1	Log2 FC LGI_T1
NHEJ1/ XLF	ENSG00000187736	All CPs vs TCL1	-1.76476	-2.54439
NR4A1	ENSG00000123358	All CPs vs TCL1	-2.27184	-2.03048
ATM	ENSG00000149311	HGI_T1vsTCL1	-0.94216	
a-EJ pathway				
Gene Symbol	ENSMBL Gene ID	Sig-DE in	Log2 FC HGI_T1	Log2 FC LGI_T1
MRE11	ENSG00000020922	All CPs vs TCL1	-0.76664	-0.73551
LIG3	ENSG00000005156	HGI_T1vsTCL1	0.77591	
POLQ	ENSG00000051341	LGI_T1vsTCL1		0.45548
XRCC1	ENSG00000073050	LGI_T1vsTCL1		0.66162

Table 6.5 DE genes associated with ALT mechanism. Genes with absolute log2 fold change > 1 shaded in yellow. FDR<0.05.

ALT related genes				
Gene Symbol	ENSMBL Gene ID	Sig-DE in	Log2 FC HGI_T1	Log2 FC LGI_T1
BLM	ENSG00000197299	All CPs vs TCL1	0.781929	0.897606
MRE11A	ENSG00000020922	All CPs vs TCL1	-0.76664	-0.73551
CBX5	ENSG00000094916	All CPs vs TCL1	0.891329	0.621386
SP100	ENSG00000067066	All CPs vs TCL1	-1.40794	-1.32716
FANCD2	ENSG00000144554	All CPs vs TCL1	0.568105	0.408381
RAD17	ENSG00000152942	All CPs vs TCL1	0.579795	0.458721
PML	ENSG00000140464	All CPs vs TCL1	-0.78523	-0.55595
RAD51D	ENSG00000185379	All CPs vs TCL1	-0.54341	-0.67854
HUS1	ENSG00000136273	HGI_T1vsTCL1	0.686401	
SLX4	ENSG00000188827	HGI_T1vsTCL1	0.735822	
ATM	ENSG00000149311	HGI_T1vsTCL1	-0.94216	
NSMCE2	ENSG00000156831	LGI_T1vsTCL1		-0.93031
FANCM	ENSG00000187790	LGI_T1vsTCL1		0.615387
BRCA2	ENSG00000139618	LGI_T1vsTCL1		0.779916
CDK2	ENSG00000123374	LGI_T1vsTCL1		-0.78406
H2AFX	ENSG00000188486	LGI_T1vsTCL1		1.061649
MND1	ENSG00000121211	LGI_T1vsTCL1		0.823339
TP53	ENSG00000141510	LGI_T1vsTCL1		0.662605
RPA1	ENSG00000132383	LGI_T1vsTCL1		0.495498
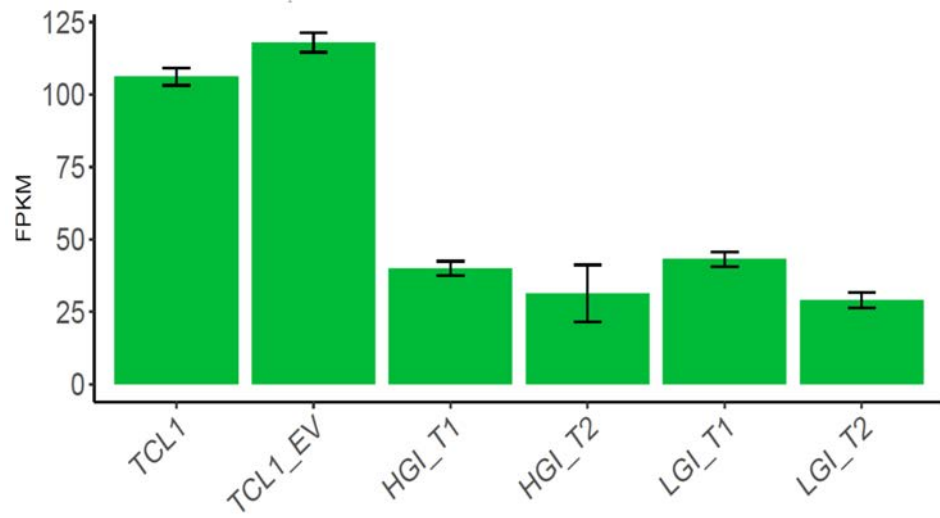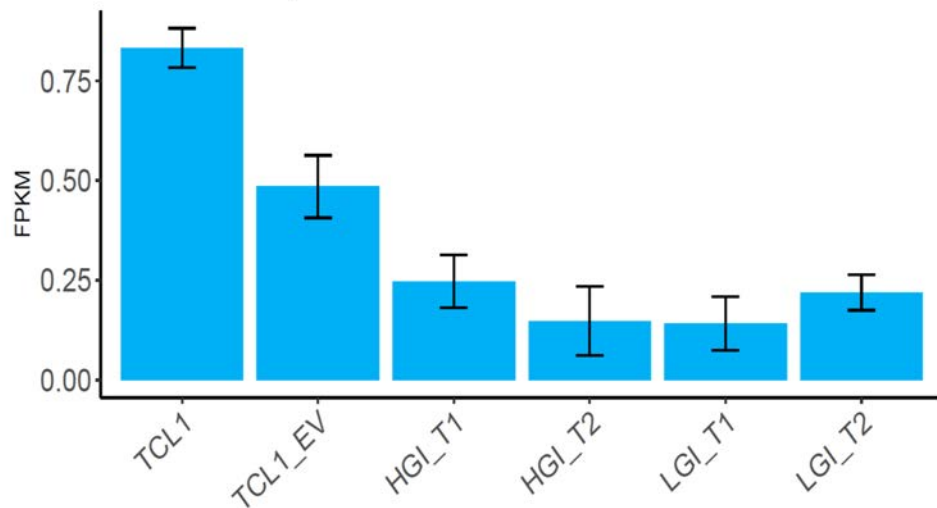
Figure 6.10 Gene expression of some ALT and NHEJ related genes in early and late post-crisis. FPKM normalisation of SP100 in (A), NHEJ1 in (B). The y-axis shows the relative gene expression represent by FPKM, while the y-axis represent the pre- and post-crisis subgroups at T1 and T2.

6.3.2 WES analysis of CCD18LU-ATRX-/- post-crisis clonal populations

6.3.2.1 Alignment and read depth assessment of WES data

In this work, we used WES to characterise genetic changes associated with the loss of ATRX and activation of ALT mechanism. DNA samples were extracted from the normal fibroblast CCD18LU cell line "Parental", pre-crisis TCL1, pre-crisis TCL1-EV, and early post-crisis (T1) were sent to Novogene for WES. For each sample, the adaptor sequences were trimmed from the FASTQ data (raw PE reads that passed the Illumina Chastity filter) using Skewer v0.2.2 (H. Jiang *et al.*, 2014). During the trimming process, the reads also were filtered with base quality score <Q20, where ~99.9% of reads were passed this step (Appendix 6.6). Overall, about 99.9% of trimmed reads were aligned to the human genome reference as paired end reads (PE) with various mapping quaintly scores, and ~96% of trimmed reads were aligned to human reference genome (human_g1k_v37_decoy, downloaded from 1000 genome) with mapping quality $\geq$ Q20 (<1/100 chance that the alignment is wrong). Among the mapped reads, 8.8-10% were marked as PCR duplicates and approximately 0.1% of reads failed to map to the human reference genome. In summary, more than 6 X 10^9 bases were aligned to the human genome, and about 97% of them were with base quantity score $\geq$ Q20.

Furthermore, the depth of coverage was assessed on certain chromosomes using the genomic intervals covered by SureSelect Human All Exon V6 r2 kit. Overall, the read depth was around 50X in across exons. However, some variation between samples and between chromosome regions were observed. For example, the analysis of read depth on chromosome 2 showed some variation in coverage that ranged between 10 and 150X with the median $\geq$ 45X (Figure 6.11).
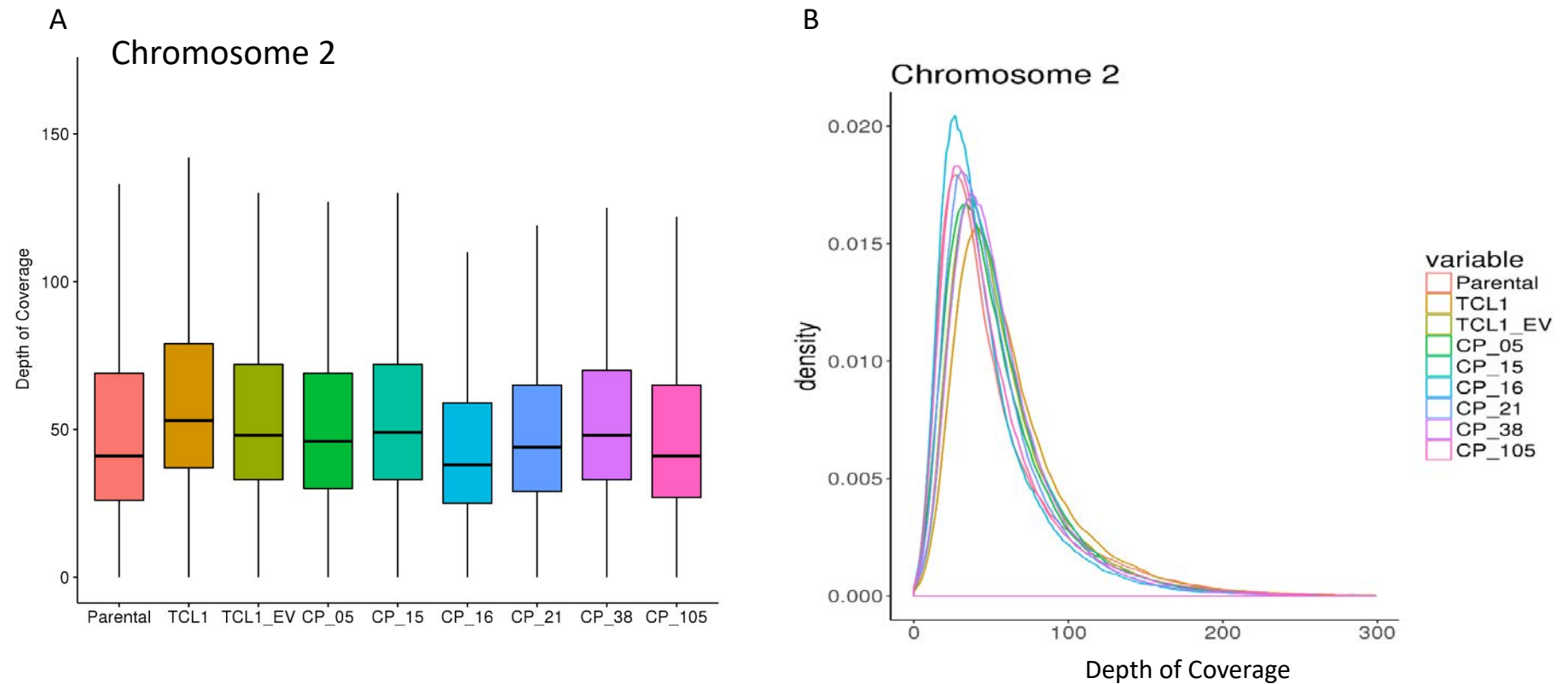
Figure 6.11 Depth coverage analysis of WES on chromosome 2. (A) Boxplots of the depth of coverage across the samples. It shows the median with lower and upper intervals. (B) The density plot for the depth of coverage showing variation between regions that ranged between 10 and 150X with the majority at 50X.

6.3.2.2 Analysis of frequently mutated genes in post-crisis CPs

Clonal populations that survived crisis showed a degree of subclonality represented by the number of colonies that emerged from the crisis. This may mimic the genetic heterogeneity seen in cancers. The study of the subclonality is important to understanding the processes underlying immortalisation and the activation of ALT. However, the genetic complexity of subclonality makes it difficult to identify the somatic Single Nucleotide Variants (SNVs). The presence of a mutation in only a small fraction of cells, due to clonal heterogeneity, makes it difficult to detect such mutations. Bioinformatics programs that identify SNVs vary in sensitivity and specificity (Kroigard *et al.*, 2016; Cai *et al.*, 2016). Therefore, three of the commonly used programs were used to analyse the WES data from the nine samples in order to identify SNVs and indels that are associated with ATRX knock-out and activation of ALT. The somatic calling of SNVs and indels was conducted as described in section in section 2.2.3.5.1.

In this work, the normal CCD18LU "parental" was used as the normal baseline for identification of somatically acquired SNVs and indels in the other DNA samples. The analysis showed a very large number of SNVs were identified by Somatic-Sniper compared to Strelka2 and Mutect2 (Figure 6.12), however >90% of SNVs identified by the Somatic Sniper software were in intronic regions with <5X coverage (Figure 6.12). Once these were removed the number of SNVs identified by the three programs were similar but more SNVs were detected in the post-crisis CPs using Strelka2. For indel analysis, the Mutect2 program was more sensitive than Strelka2. In the pre-crisis samples 1-2 indels were detected rising to 4-10 in the post-crisis CPs. Visualisation of these indels in the Bam files using IGV showed that some of them were in repetitive regions suggesting false positive detection as a results of mapping problem at these regions.
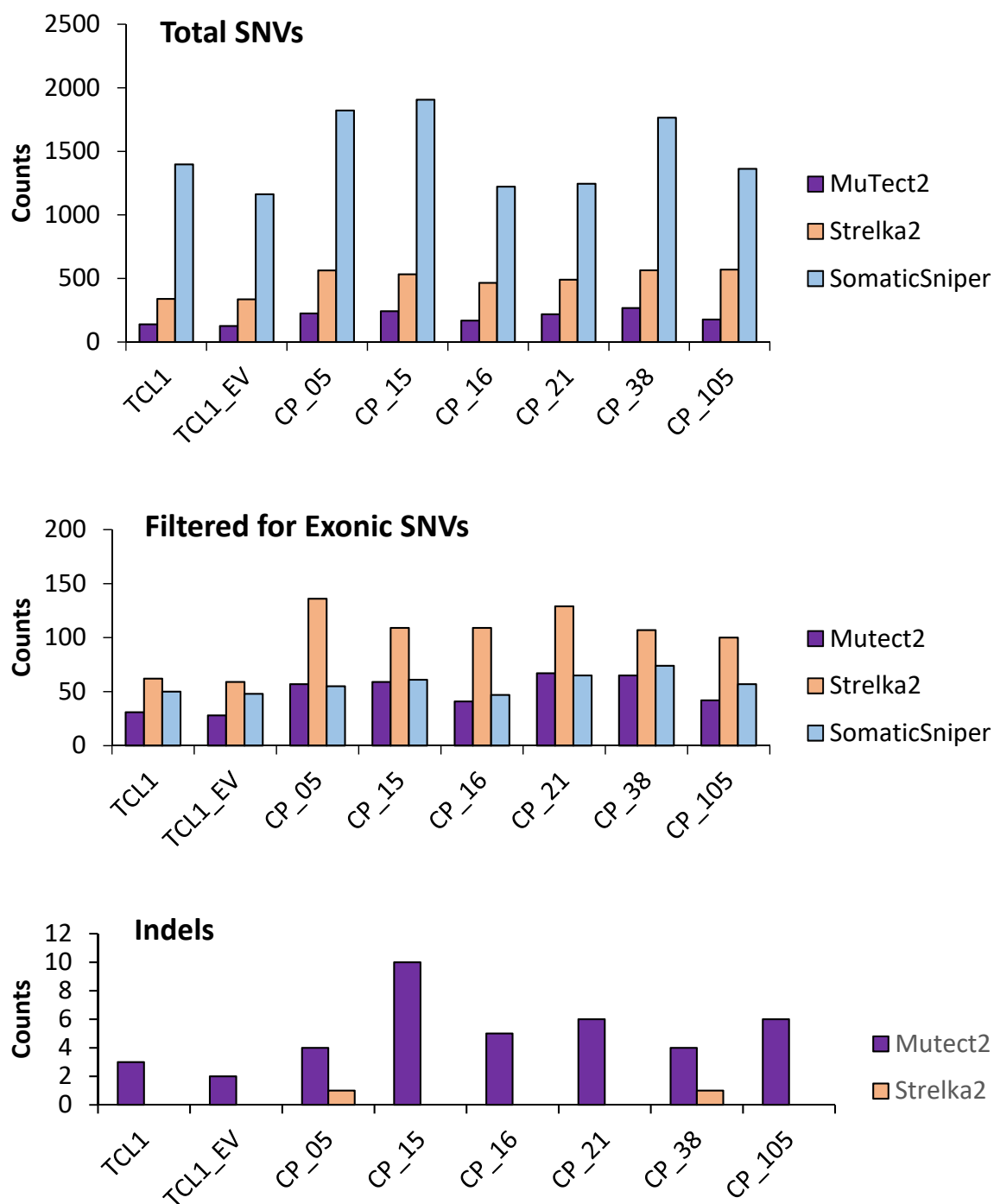
Figure 6.12 Comparison of the number SNVs and indels detected using various somatic callers. Top graph shows the total number of SNVs reported by each program. Middle graph shows the exonic SNVs detected by each tool. Lower graph shows the total indels detected by Mutect2 and Strelka2

The overlap between somatic SNV calls showed that low percentage (~ 13.5%) of somatic SNV calls were overlap between these three somatic callers (Figure 6.13). Therefore, only the SNVs that were detected by at least two somatic callers were included in the downstream analyses.
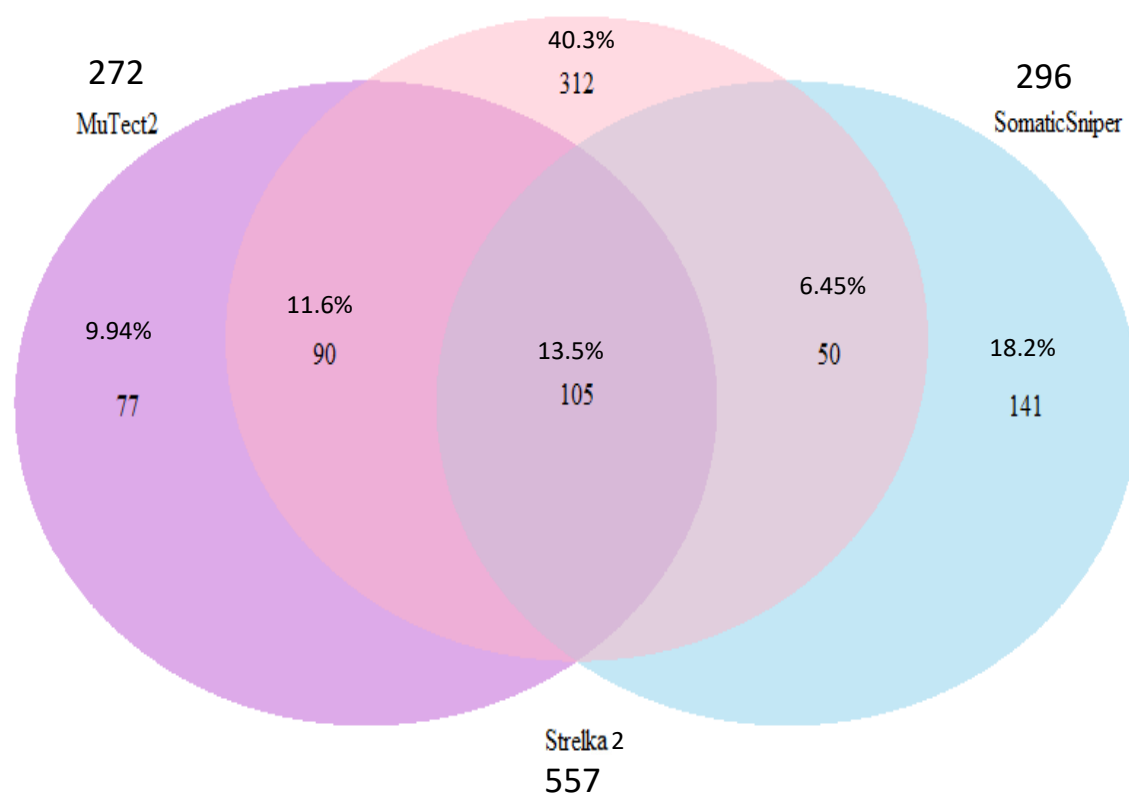
Figure 6.13 Overlap in somatic SNV calls. Somatic mutation calls in Exonic+UTRs regions were compared across three callers (Mutect 2, Strelka 2, SomaticSniper) in six samples of crisis surviving CPs. The parental CCD18LU was used as the baseline for somatic calling.

The somatic SNVs was combined with the somatic indels detected with Mutect2 (Table 6.6). Annotation of non-synonymous SNVs and frameshift indels showed that the majority of detected mutations are predicted to give rise to a missense mutation in the coding sequence. In comparison to the parental CCD18LU cells, the post-crisis clonal populations accumulated more missense mutations than the pre-crisis TCL1. The number of the missense mutations range was 24-35 in CP_05, CP_15, CP_16, and 36-47 in CP_21, CP_38, CP_105. In particular, there was a remarkable number of missense mutations in CP_21.

Table 6.6 Frequency analysis of exonic mutations in the pre- and post-crisis samples. The parental CCD18LU was used as the baseline.

Mutation type	TCL1	TCL1_EV	CP_05	CP_15	CP_16	CP_21	CP_38	CP_105
Missense	21	21	34	35	24	47	36	37
Frame_Shift_Del	2	1	4	3	2	6	2	2
Frame_Shift_Ins	3	1	1	3	3	2	3	2
In_Frame_Del	1	1	1	0	1	2	1	1
Nonsense	0	0	0	0	1	1	2	1
Splice_Site	1	1	1	1	3	1	4	2
In_Frame_Ins	0	1	1	0	0	0	0	1
Total	28	26	42	42	34	59	48	46
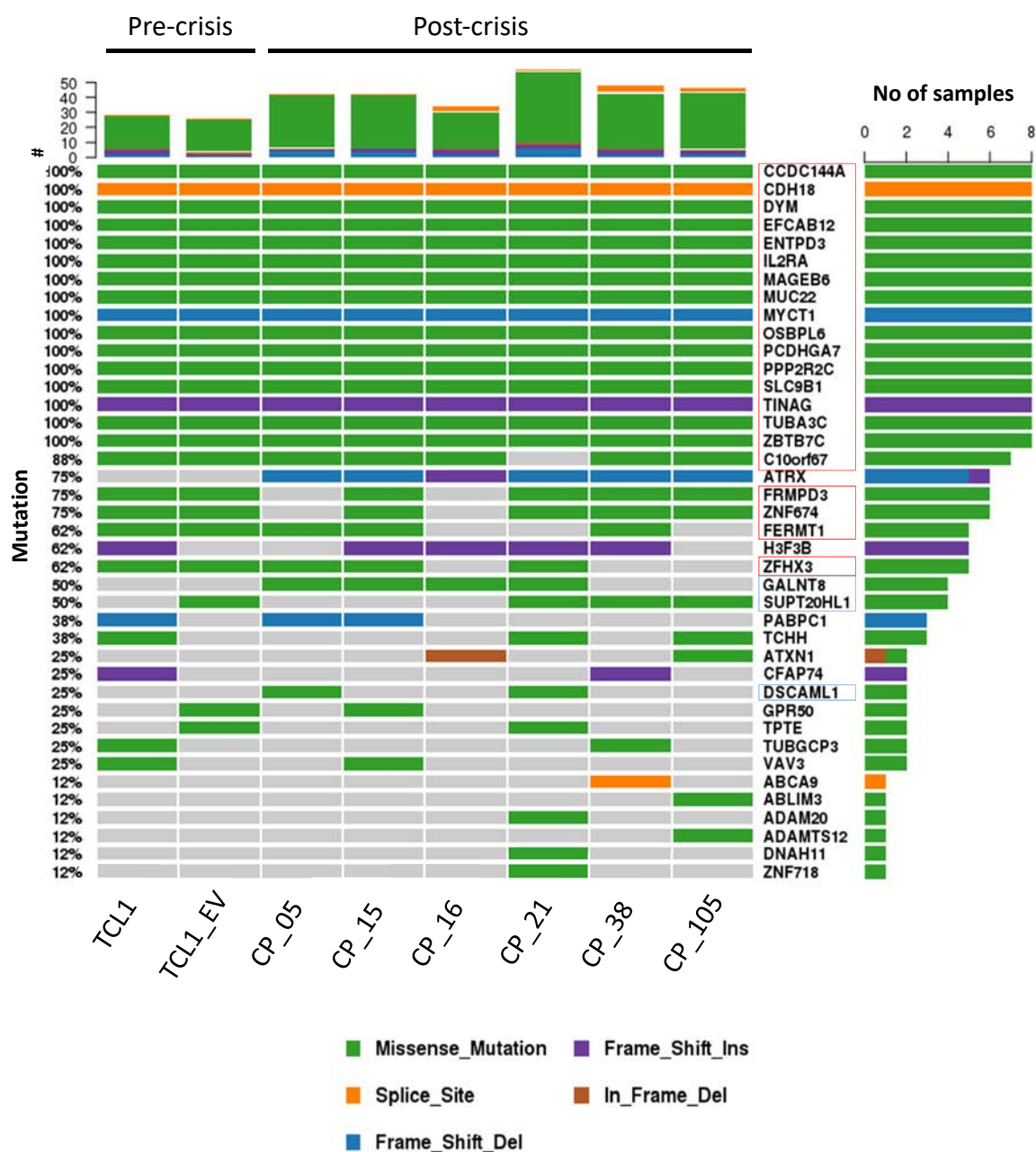
Figure 6.14 Oncoplot of somatic non-synonymous SNVs and indels detected in pre-crisis and post-crisis CPs. The red box highlights the genes that are mutated in TCL1 and consequently in the post-crisis CPs. The blue box highlights ambiguous somatic SNVs detected in post-crisis CPs

In addition to the ATRX mutation analysis that have discussed in the chapter 5, analysis of the most frequently mutated genes showed that sixteen genes are mutated in 100% (8/8) of the samples including the pre-crisis TCL1, and five genes were mutated in 75% (6/8) of the samples also included TCL1 (Figure 6.14). Genes were mutated at a single locus that also found in TCL1, which suggested these mutations occurred during the pre-crisis SV40-transformation in TCL1, and thereby inherited to the post-crisis CPs. The functional annotation analysis of SNV and indel mutations (Table 6.7) showed that mutations in ten genes (*PCDHGA7, ZBTB7C, OSBPL6, PPP2R2C, DYM, OSBPL6, C10orf67, CCDC144A, ZFHX3, FRMPD3*) were predicted to cause protein damage, and another mutations found in eight genes (*IL2RA, EFCAB12, ENTPD3, SLC9B1, MUC22, MAGEB6, FERMT1, ZNF674*) were predicted to be tolerated, while the consequences of mutations in four genes (*CDH18, TUBA3C, MYCT1, TINAG*) were not predictable. The IGV visualisation of aligned reads bam files has confirmed the results from somatic calling in these genes. For example, the SNV mutation (chr13:19751700 A>T) in *TUBA3C* seems to be exist in 26% (35/137) of reads in TCL1, and 52% (39/75) of reads in CP_05, while 0% (0/115) of reads in parental CCD18LU (Figure 6.15).

Gene Symbol	Chr	Start	End	Ref	Alt	Func. refGene	ExonicFunc. refGene	SIFT score	SIFT_converted Rank score	SIFT Prediction
C10orf67	10	23633648	23633648	C	G	exonic	nonsynonymous SNV	0	0.912	Deleterious
CCDC144A	17	16593976	16593976	G	T	exonic	nonsynonymous SNV	0	0.912	Deleterious
ZBTB7C	18	45566424	45566424	A	G	exonic	nonsynonymous SNV	0.022	0.486	Deleterious
DYM	18	46690113	46690113	T	A	exonic	nonsynonymous SNV	0.007	0.599	Deleterious
OSBPL6	2	179251800	179251800	A	C	exonic	nonsynonymous SNV	0.001	0.784	Deleterious
PPP2R2C	4	6382763	6382763	G	C	exonic	nonsynonymous SNV	0	0.912	Deleterious
PCDHGA7	5	140762559	140762559	T	G	exonic	nonsynonymous SNV	0.018	0.506	Deleterious
ZFHX3	16	72829974	72829974	G	A	exonic	nonsynonymous SNV	0	0.912	Deleterious
FRMPD3	X	106844526	106844526	G	A	exonic	nonsynonymous SNV	0.002	0.721	Deleterious
IL2RA	10	6104063	6104063	C	A	exonic	nonsynonymous SNV	0.052	0.39	Tolerated
EFCAB12	3	129130150	129130150	T	G	exonic	nonsynonymous SNV	0.175	0.491	Tolerated
ENTPD3	3	40464465	40464465	C	G	exonic	nonsynonymous SNV	0.071	0.784	Tolerated
SLC9B1	4	103853410	103853410	T	G	exonic	nonsynonymous SNV	1	0.01	Tolerated
MUC22	6	30994134	30994134	T	C	exonic	nonsynonymous SNV	1	0.01	Tolerated
MAGEB6	X	26212379	26212379	C	T	exonic	nonsynonymous SNV	0.175	0.224	Tolerated
FERMT1	20	6093177	6093177	A	G	exonic	nonsynonymous SNV	0.344	0.125	Tolerated
ZNF674	X	46359922	46359922	C	T	exonic	nonsynonymous SNV	0.156	0.241	Tolerated
CDH18	5	19838867	19838867	C	T	splicing	.	.	.	N/A
TUBA3C	13	19751700	19751700	A	T	exonic	nonsynonymous SNV	.	.	N/A
MYCT1	6	153019100	153019103	AGAT	-	exonic	frameshift deletion	.	.	N/A
TINAG	6	54254603	54254603	-	TGGG	exonic	frameshift insertion	.	.	N/A

Table 6.7 Annotation analysis of frequency mutated genes in TCL1 and post-crisis CPs. SIFT predicts the possible effects of an amino acid substitution on protein function. The SIFT predication outcomes include deleterious (protein damaging), tolerated, N/A (no prediction).
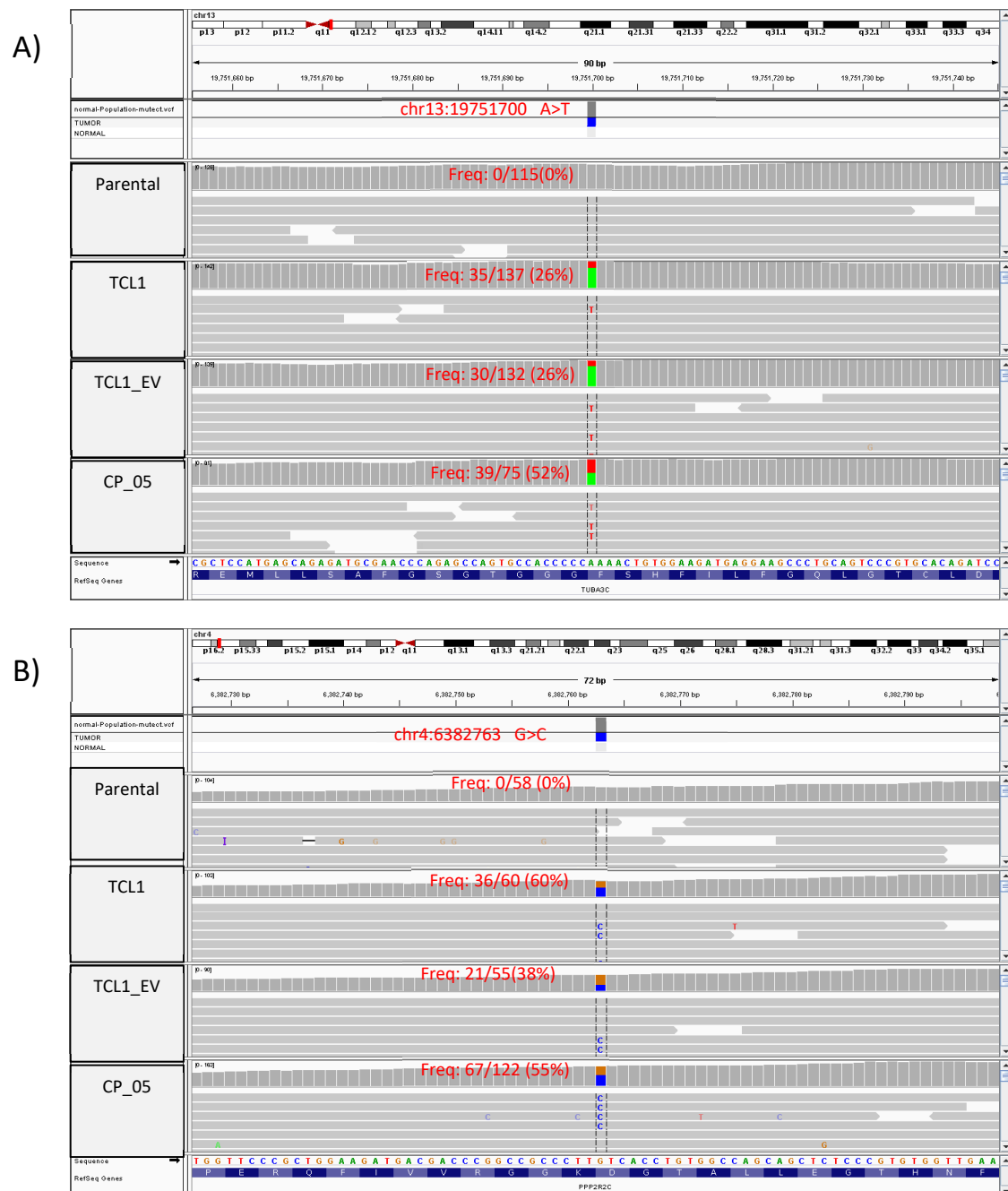
Figure 6.15 Visual inspection of some validated somatic SNVs in post-crisis CPs. (A) Chr13:19751700 A>T in *TUBA3C*. (B) Chr4:6382763 G>C in *PPP2R2C*.

Three somatic SNVs (one SNV per gene) were detected in *GALNT8, DSCAML1*, and *SUPT20HL1* in some of post-crisis CPs but not in TCL1 (Table 6.8). The SNVs in *GALNT8, DSCAML1* were predicted a protein damaging, while unknown consequence of mutation in *SUPT20HL1* was predicted. However, the IGV visualisation of aligned reads bam files showed sort of uncertainty in the detection of the SNVs in these three genes (Figure 6.16).

The ChrX:24382453   G>C in *SUPT20HL1* was in a repetitive region where mapping error could results in false positive detection. Another example, the detected SNV mutation (Chr12:4874692 A>G) in *GALNT8* seems to be associated with changes in the allele frequency, as it showed in 50% (11/22) of reads in TCL1, and 75% (18/24) of reads in CP_05, however, the read depth was only 8X in parental locus which may explain why the somatic caller has failed to detected in TCL1. Furthermore, false positive detection of indel mutations (using Mutect2) was observed in H2F3B (Figure 6.17) and some other genes.

Gene.refGene	SNVs detected in	Chr	Start	End	Ref	Alt	Func.refGene	ExonicFunc. refGene	SIFT score	SIFT_converted Rank score	SIFT Prediction
GALNT8	CP_05 CP_15 CP_16 CP_21	12	4874692	4874692	A	G	exonic	nonsynonymous SNV	0.014	0.531	Deleterious
DSCAML1	CP_05 CP_21	11	117302380	117302380	C	A	exonic	nonsynonymous SNV	0.004	0.682	Deleterious
SUPT20HL1	CP_21 CP_38 CP_105	X	24382453	24382453	G	C	exonic	nonsynonymous SNV	.	.	N/A

Table 6.8 Annotation analysis of ambiguous somatic SNVs detected in post-crisis CPs. SIFT predicts the possible effects of an amino acid substitution on protein function. The SIFT predication outcomes include deleterious (protein damaging), tolerated, N/A (no prediction)

Figure 6.16 Ambiguous somatic SNVs in some post-crisis CPs (A) Chr12:4874692 A>G in *GALNT8*. (B) ChrX:24382453   G>C in *SUPT20HL1*.
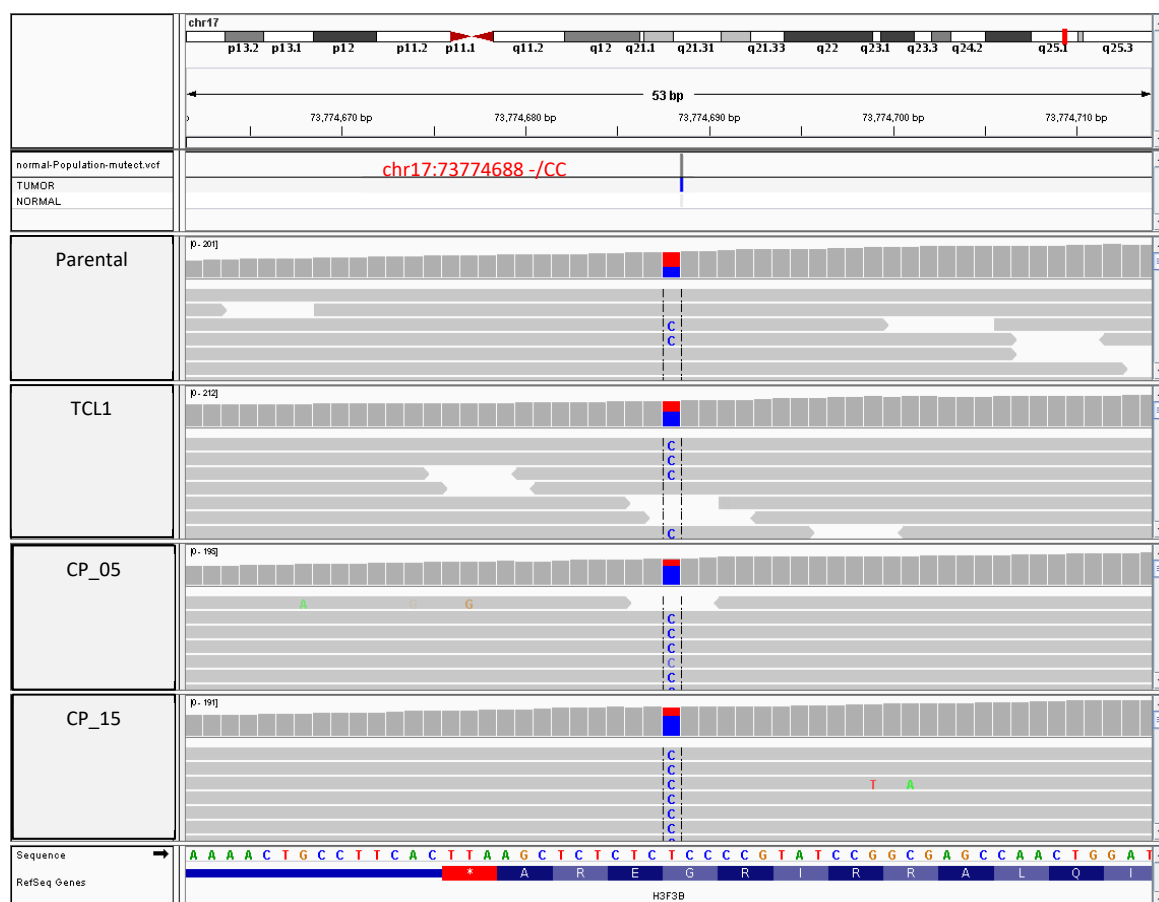
Figure 6.17 False postive somatic frameshift insertion detected in *H3F3B* in TCL1 and post-crisis CPs.

The effects that associated with inactivating mutations in commonly mutated genes in TCL1 and in post-crisis CPs have been discussed here. *PCDHGA7* (Protocadherin Gamma Subfamily A 7) is a neural cadherin-like cell adhesion gene and codes for a potential calcium-dependent cell-adhesion protein (Wu and Maniatis, 1999). *CDH18* (Cadherin 18) is a type II classical cadherin gene coding for a protein associated with calcium-dependent cell-cell adhesion (Shibata *et al.*, 1997). Cadherin and protocadherin genes are widely mutated or down regulated in carcinomas and epithelial cells transformed with SV40-LT (Gunduz *et al.*, 2012). Absence of cadherin/ protocadherin proteins is implicated in loss of cell-extracellular matrix adhesion, which promotes cellular growth and metastasis of tumours (Navarro *et al.*, 1991; Moh and Shen, 2009). *ZBTB7C* (Zinc Finger and BTB Domain Containing 7C) encodes for a predicted cell growth inhibitor protein. It was found that ectopic expression of *ZBTB7C* results in

significant reduction in cell growth of cervical carcinoma-derived cell lines such as HeLa and CaSki (Reuter *et al.*, 1998). *OSBPL6* (Oxysterol Binding Protein Like 6) encodes for a member OxySterol Binding Protein (OSBP) family of intracellular lipid receptors (Jaworski *et al.*, 2001). Although the role of OSBPL6 in cancer is unclear, but recently there is an increased number of studies about the role of lipid metabolism in tumorigenesis by regulating certain pathways in cancers. *PPP2R2C* (Protein Phosphatase 2 Regulatory Subunit B gamma) encodes for one of the four major Ser/Thr phosphatases. They act upon phosphorylated serine/threonine residues in proteins results in regulation of proteins, thereby cell signalling (Shi, 2009).

ZFHX3 (Zinc Finger Homeobox 3) encodes for a transcription factor that can act as an activator or a repressor. One of its main pathways upregulates CDKN1A/p21 promoter activity following TGF-beta stimulation (Mabuchi *et al.*, 2010). Thus loss of ZFHX3 seems to be associated with avoiding the growth suppression by the TGF-beta signalling pathway. *TUBA3C* (Tubulin Alpha 3c) is one of several genes coding for alpha-tubulin in human, which is the major subunit of microtubules. Alpha- and beta- tubulins equally participate in making the protofilaments of a microtubule (Nogales, 2000). Microtubules form the mitotic spindle which controls chromosomal segregation during cell division. Thus, mutations in tubulin result in microtubule dysregulation that contributes to several tumour markers, especially the genomic instability and tetraploidy (Hernandez and Tirnauer, 2010). Furthermore, it was found that microtubules are involved in regulation of gene expression (Rosette and Karin, 1995).

FRMPD3 (FERM and PDZ Domain Containing 3) encodes for a protein that contains a protein-protein interaction domain called the PDZ domain. FRMPD3 is localised to the plasma membrane and is associated with the cytoskeleton, thus it predicted to play a role in the innate immune response by interaction with some Toll-like receptors network (Hung *et al.*, 2018). *DYM* (Dymeclin) encodes a peripheral membrane protein which shuttles between the cytosol and Golgi complex. It is believed to play a role in the regulation of Golgi-associated secretion. Mutations in DYM gene are associated with an autosomal recessive disorder called Dyggve-Melchior-Clausen (DMC) (El Ghouzzi *et al.*, 2003). However, there is no evidence that DYM actually plays a role in cancers, which suggests that the DYM mutation in TCL1 pre-crisis was a random mutation as a result of inactivation of the M1 of the checkpoint by SV40-LT.

6.3.2.3 Analysis of CNVs and genomic instability in post-crisis CPs

To investigate the telomere driven genomic instability associated with the emergence from the crisis and activation of ALT, the allele-specific CNV profiles of post-crisis CPs was conducted using WES data as described in section 2.2.3.51. This analysis included detecting Loss Of Heterozygosity (LOH) and estimating the ploidy in comparison to the CCD18LU parental cells.

The TCL1 and TCL_EV cell populations showed relatively stable diploid genomes but with a limited number of regions that showed LOH in chromosomes 8, 11, 18, 19 with estimated cellular fraction 82%, 57%, 63% and 48% respectively (Figure 6.18 and Appendix 6.7). However, the early post-crisis CPs had an estimated ploidy of three and a high level of genomic instability. Segmental amplifications were predicted in chromosome 8 and 19 in CP_05, and in chromosome 19 in CP_38 (Table 6.9 and Appendix 6.8). The annotation of amplified segments in chromosome 19 showed a minimum overlap between the CP_05, and CP_38 chr19 that involved two zinc finger genes (*ZNF99*, *ZNF492*), which may play roles the regulation of gene expression required to escape crisis.

Table 6.9 Amplified chromosomal segments in the post-crisis CPs. Amplification was calculated (2 * average of ploidy +1).

Sample ID	chrom	start	end	Estimated cellular fraction for segment with amplification	Estimated total copy number	Estimated minor allele copy number
CP_05	8	116699	1497587	0.911318	9	1
CP_05	19	22654436	22940507	0.911318	13	1
CP_38	19	20807847	24310803	0.912882	9	2
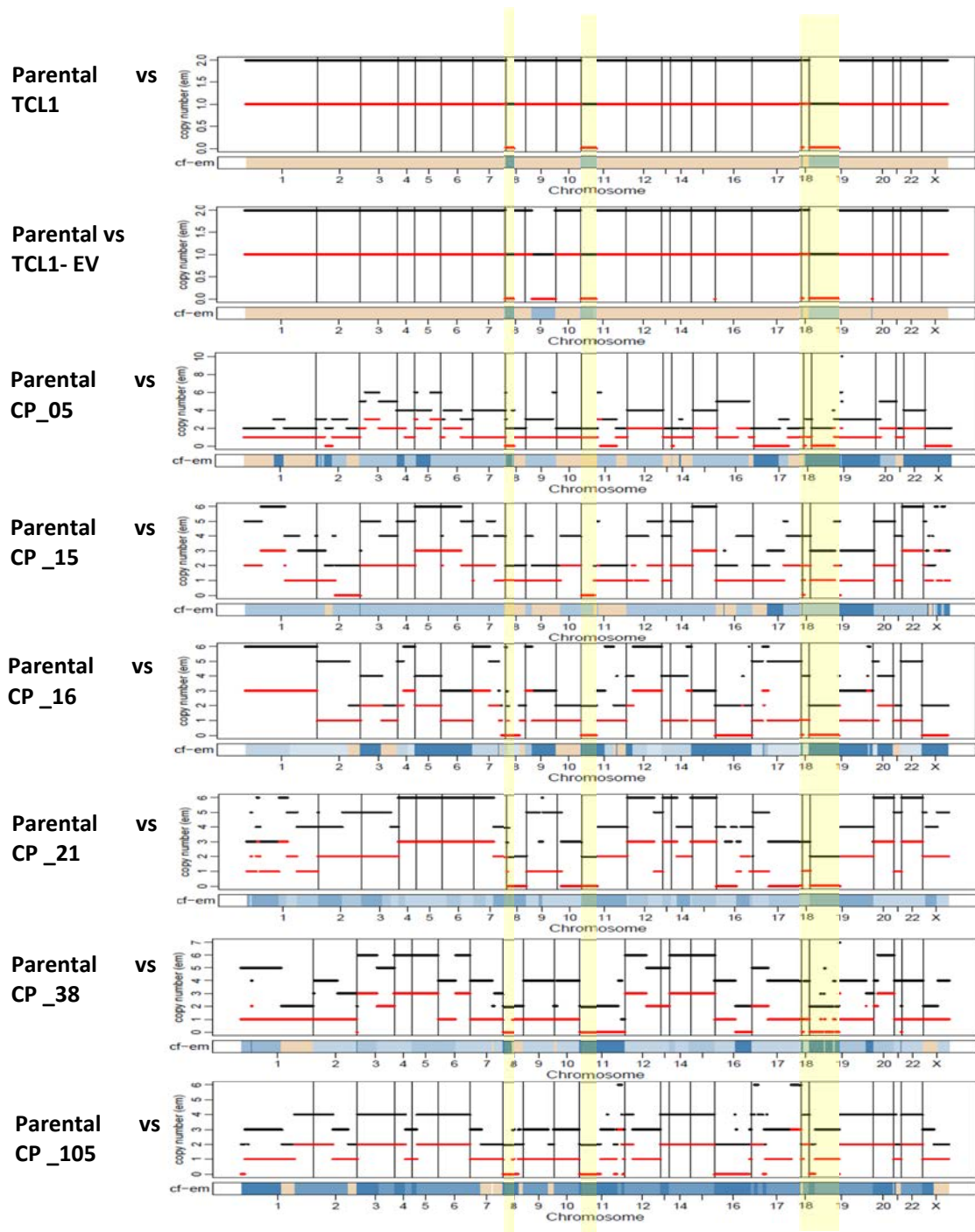
Figure 6.18 Genome view of CNV analysis. The CCD18LU parental cell line was used as the baseline. The plots show total copy number for each segment (black) and 'minor allele' copy number for each segment (red). The bottom bar shows the associated cellular fraction (cf). Dark blue indicates high cf. Light blue indicates low cf. Beige indicates a normal segment (total=2,minor=1). Regions with LOH in TCL1 are shaded in yellow.

The estimated cellular fractions varied from 52% to 93% between the post-crisis CPs and showed some relationship to the estimated number of subclones contributing to each CP that emerged from crisis (Table 6.10). The estimated ploidy predicts that some cells in each CP have undergone tetraploidization. As expected the post-crisis CPs showed the LOHs in chromosomes 8, 11, 18, 19 that were seen in TCL1 but with some additional variations (Table 6.10). For example the TCL1 chromosome 11 LOH was observed in all post-crisis CPs except in CP_05 which showed the LOH in a different region of chromosome 11. The LOH in chromosome 8 was detected in all post-crisis CPs except CP_15 and the LOH in chromosome18 was also present in all post-crisis CPs except CP_21, and CP_105.

Table 6.10 Summary of CNV analysis in post-crisis CPs.

Sample ID	Estimated Ploidy	Cellularity estimated	Estimated No. of subclones emerging from crisis	Chromosomes with segments showing LOH	Chromosomes with amplified segments
TCL1	1.92	0.58	-	8, 11, 18, 19	
TCL1_EV	1.8	0.46	-	8, 9, 11, 18, 19	
CP_05	2.8	0.52	7	2, 8, 11, 17, 18, X	8, 19
CP_15	2.74	0.74	12	2, 11, 18, 19	
CP_16	2.98	0.92	8	7, 8, 11, 16, 18, X	
CP_21	4	0.57	4	8, 10, 11, 16, 17, 19	
CP_38	2.94	0.88	1	8, 11, 16, 18, 21	19
CP_105	3.16	0.93	2	1, 8, 11, 16, 17, 19	

To verify the CNV results, the read depths at heterozygous SNPs present in the CCD18LU parental cell line (germline SNPs) were visualised using IGV (Figure 6.19). For example, for the germline SNP at Chr8:41585438 G>A, both the G and A alleles were observed equally in the parental cells, while the number of the aligned reads that contained A was reduced to 20% in TCL1 and CP_05, and to 0% in CP_38, CP_21, and CP_105, while the ratio of reads that contained A was approximately 50% in CP_15, and CP_16. The loss of heterozygosity seen in several regions in the pre-crisis TCL1 and TCL_EV may be associated with uncontrolled cellular proliferation as a result of p53 and Rb inactivation by SV40-LT. However, the increase in average ploidy and additional changes in allelic frequencies in post-crisis CPs may arise from the telomere break-fusion-bridge cycles and tetraploidization events during the crisis. There may also be some alterations to the ratios between the subclonal populations within each post-crisis CP that contributes to the observed changes in allele frequencies.

In addition, the different post-crisis CPs showed similar LOH in some chromosomal regions that may participate in similar changes in gene expression. The CP_05, and CP_15 showed LOH in chromosome 2 which partially separates them from the other the CPs. On the other hand there was LOH in similar regions of chromosome 16 in CP_16, CP_21, CP_38, CP_105, but not in CP_05, and CP_15. These results may support the clustering of post-crisis samples base on gene expression.
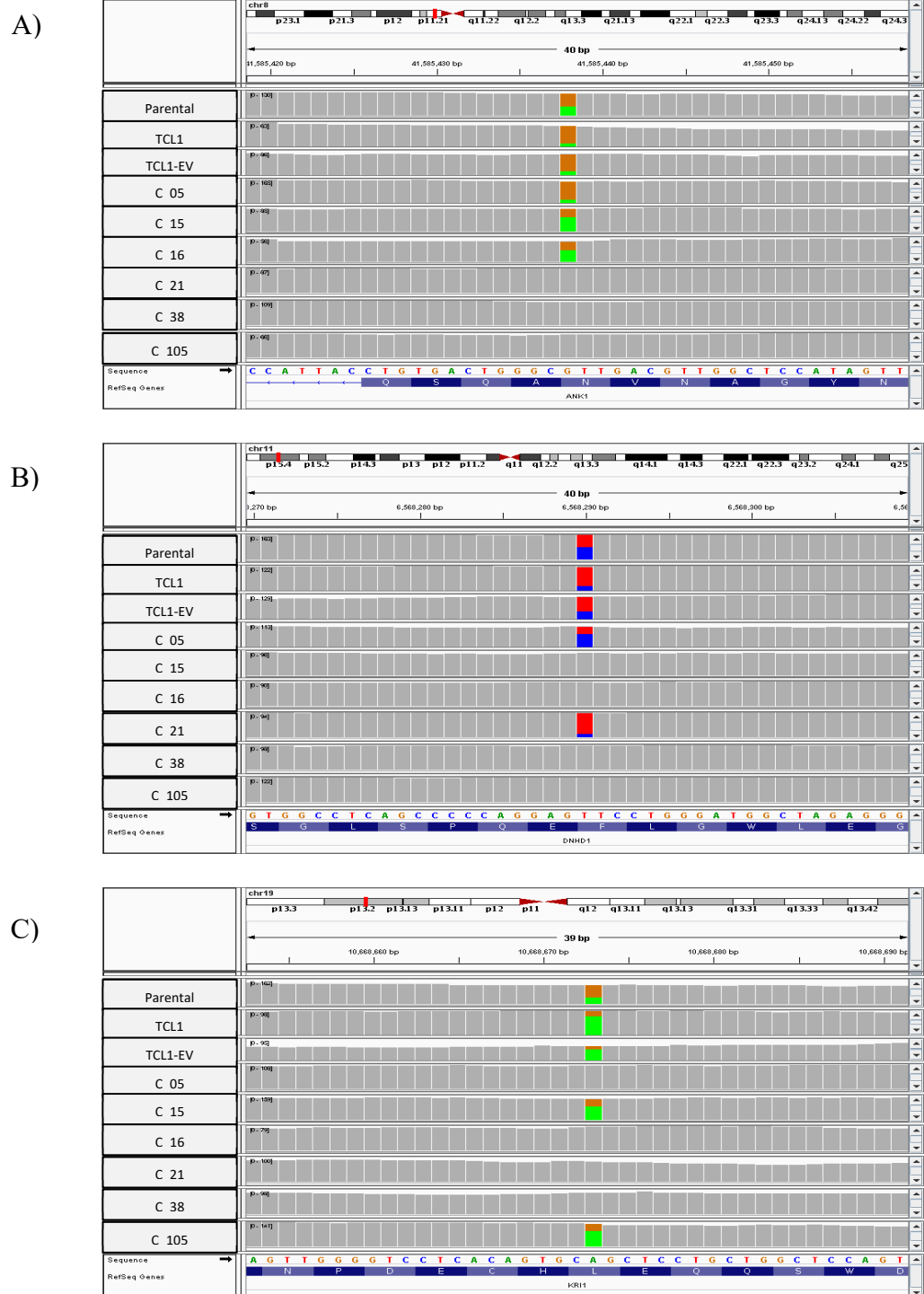
Figure 6.19 Visual inspection of SNPs that are heterozygous in the CCD18LU parental cell line in some chromosomal regions in chr8, chr11, chr19. (A) Chr8:41585438 G>A in *ANK1*. (B) Chr11:6568290 T>C in *DNHD1*. (C) Chr19:10,668,670 A>G in *KRI1*.

To characterise the LOH in TCL1, the list of genes related to each segment with LOH were retrieved from UCSC genome browser (Table 6.11). The chromosome 8 segment that showed LOH in 80% of TCL1 cells mainly contained genes that are associated with regulation of protein phosphorylation such as (*PPP1R3A, PPP2CB, and PPP3CC*). LOH in the chromosome 11 segment (57% of TCL1 cells) included genes associated with the innate immune response and type I interferons such Interferon Induced Transmembrane Proteins (*IFITM1, IFITM2, IFITM 3, IFITM 5, IFITM 10*), and Interferon Regulatory Factor 7 (*INF7*), as well as large set of Tripartite Motif Containing genes (*TRIM3, TRIM5, TRIM6, TRIM21, TRIM22, TRIM66, TRIM68*). LOH in the chromosome 18 segment detected in 63% of TCL1 cells included genes that are associated with focal adhesion such as Laminin Subunit Alpha 1 (*LAMA1*), which is part of extracellular matrix glycoproteins. The chromosome 18 segment showing LOH in 48% of TCL1 cells included 81 genes that encodes zinc-finger proteins associated with regulation of gene expression, a member of the Janus kinase family gene (*JAK3*), as well as two members of MAP kinase gene family (*MAP2K2, MAP2K7*), a member of Phospholipid Phosphatase 2 (*PPAP2C*), a member of Phosphatidylinositol 3-kinase (*PIK3R2*) and SWI/SNF related, Matrix associated, Actin dependent Regulator of Chromatin (*SMARCA4*).

Table 6.11 List of chromosomal segments with LOH in TCL1.

Chrom	Start	End	cellular fraction	Total CNV	B-allele CNV	Segment size (kb)	Number of effected genes
chr8	163635	43152208	0.815716	1	0	42988.57	372
chr11	87042	51412083	0.574105	1	0	51325.04	701
Chr18	47692	15271065	0.635019	1	0	15223.37	327
chr19	110812	22940046	0.482675	1	0	22829.23	761

Functional analysis of the combined list of 1,434 genes associated with LOH in TCL1 (in chr8, ch11, ch18, and ch19) was investigated using DAVID bioinformatics resource (Table 6.12). The GO term enrichment analysis of biological processes showed 122 genes were involved in G-protein coupled receptor signalling pathway and 177 genes were associated with regulation of gene expression. The enrichment of GO terms associated with cellular components showed 384 genes in nucleus and 366 genes were associated with terms related to integral components of the cell membrane. The enrichment of GO terms associated with molecular functions revealed eight genes related to methyl-CpG binding. These include DNA methyltransferase 1(DNMT1), methyl-CpG binding domain protein 3 (MBD3), five members of methyl-CpG binding domain protein 3 like (MBD3L1-5), and a ubiquitin-like with PHD and ring finger domains 1(UHRF1).

Table 6.12 Enrichment analysis of GO terms associated with the 1434 genes that showed LOH in TCL1. BP= biological process, CC= cellular components, MF= molecular functions.

GO term	Description	Gene ratio	%	p value	p adj	Counts
BP	detection of chemical stimulus involved in sensory perception of smell	95/1434	6.6	2.10E-26	6.60E-23	95
BP	G-protein coupled receptor signalling pathway	122/1434	8.5	1.50E-14	2.30E-11	122
BP	signal transduction	127/1434	8.9	9.70E-09	1.00E-05	127
BP	regulation of transcription, DNA-templated	150/1434	10.5	1.50E-07	1.20E-04	150
BP	transcription, DNA-templated	177/1434	12.3	5.70E-06	3.50E-03	177
BP	sensory perception of smell	26/1434	1.8	4.10E-05	2.10E-02	26
CC	intracellular	131/1434	9.1	9.90E-07	5.60E-04	131
CC	Golgi lumen	18/1434	1.3	1.30E-04	3.50E-02	18
CC	nucleus	384/1434	26.8	1.50E-02	8.80E-01	384
CC	plasma membrane	297/1434	20.7	1.80E-02	8.70E-01	297
CC	integral component of membrane	366/1434	25.5	1.90E-02	8.30E-01	366

MF	olfactory receptor activity	95/1434	6.6	2.50E-26	2.50E-23	95
MF	G-protein coupled receptor activity	110/1434	7.7	9.10E-18	4.50E-15	110
MF	nucleic acid binding	107/1434	7.5	1.30E-07	4.40E-05	107
MF	RNA polymerase II core promoter proximal region sequence-specific DNA binding	45/1434	3.1	2.70E-05	6.60E-03	45
MF	methyl-CpG binding	8/1434	0.6	2.50E-04	4.80E-02	8
MF	metal ion binding	171/1434	11.9	5.50E-04	8.70E-02	171

6.4 Discussion

6.4.1 High level of genomic instability associated with activation of ALT

The data presented in the previous chapter showed that introducing inactivating mutations into *ATRX* in a pre-crisis clone of SV40-transformed fibroblasts facilitated and increased the probability of ALT activation in crisis-surviving clonal populations. Despite the presence of wild type *TERT*, all the *ATRX*[-/-] post-crisis CPs had activated ALT. In this chapter, integrated analyses of WES data and transcriptome profiling (RNA-seq) were conducted to characterise the genetic and transcriptional alterations that may be required for activation of ALT at or following crisis. At the early post-crisis time point, the CPs clustered in two subgroups CP_05, CP_15, CP_16 in subgroup A, and CP_21, CP_38, CP_105 in subgroup B. The CPs in subgroup A were associated with relatively high level of micronuclei (23- 26%) in comparison to CPs in subgroup B (15-16%). Micronuclei formation may have been associated with the predicted tetraploidization events around crisis and the higher frequency at T1 supports the proposal that some of the cells were closer to crisis at this stage. The micronucleation phenotype in ALT+ post-crisis CPs is likely to be partly due to the loss ATRX, and is consistent with the previous observation that lobulated nuclei formed after ATRX depletion in HeLa (TEL+) cells (Ritchie *et al.*, 2008). The high frequency of micronucleation in ALT+ post-crisis CPs is an indicative for ALT activity (Crasta *et al.*, 2012; Lovejoy *et al.*, 2012).

The CNV analysis of WES data showed that TCL1 and TCL_EV had relatively stable genomes with mainly diploid cells. In contrast, the average ploidy was three in the post-crisis CPs and associated with high genomic instability, which varied between the post-crisis CPs. These findings suggested the increase in average ploidy and additional changes in allelic frequencies in post-crisis CPs may arise from the telomere break-fusion-bridge cycles and tetraploidization events during the crisis. There may also be some alterations to the ratios between the subclonal populations within each post-crisis CP that contributes to the observed changes in allele frequencies.

6.4.2 The combined genetic and transcriptional alternations contributed to the activation of ALT.

The DE analysis of HGI_T1 and LGI_T1 against the TCL1 showed that the most striking changes involved the downregulation in expression of genes related to cell innate immunity type I interferons response to viruses, G-protein coupled receptor binding, glycoprotein binding receptors, and cytokine receptors. The downregulation of the JAK/STAT signalling pathway may facilitate the crisis by suppression of apoptotic genes but could also include other pathways that required the JAKs and the transmembrane kinase activity, such as PI3K/AKT signalling pathway.

The integration between the CNV analysis of WES and the DE analysis of RNA-seq showed the reduction in innate immune response and type I interferons type alpha such Interferon Induced Transmembrane Proteins (*IFITM1*/2/3), and Interferon Regulatory Factor 7 (*IRF7*), as well as large set of Tripartite Motif family proteins (*TRIMs*) are associated with LOH in chromosome 11. The LOH in chromosome 11 segment which predicted in 57% of TCL1 cells, ~ 53% in CP_05, CP_15, CP_16 cells, ~70% in CP_21 cells, while it is predicted to be more than 90% in CP_38, and CP_105. Furthermore, downregulation of genes that related to G-protein coupled receptor binding that may associated in downregulation of JAK/STAT at the early post-crisis was also associated with LOH in chromosome 19 (Table 6.11). This could suggest that the significant downregulation in JAK/STAT signalling pathway was to large extend a results of LOH in chromosome 11 and chromosome 19. However, the downregulation in JAK/STAT signalling pathway may be also associated with changes in gene expression following ATRX knockout.

SP100 and a set of PML nuclear body-associated genes are characterised for containing ISREs in their promoter regions, making them the primary targets of STAT1 (Stadler *et al.*, 1995; Grotzinger *et al.*, 1996; Gongora *et al.*, 1997). This could suggest that reduction in STAT1 expression could mediate the activation of ALT by suppression the expression of SP100. In addition, the significant downregulation of *NR4A1* and *NHEJ1*/ *XLF* could modulate the switch from c-NHEJ-mediated double strand break repair between telomeres of different chromosomes to telomere resection and a-NHEJ-mediated fusion of sister- telomeres as an intermediate steps towards ALT.

In summary, from the RNA-Seq and WES data analysis I present a provisional model for the activation of ALT following SV40-transformation (Figure 6.20). This model proposes that downregulation of *NR4A1* and *NHEJ1/ XLF* gene expression results in a reduction of DSBs repair by c-NHEJ and thereby a reduction in inter-chromosomal fusion events. This combined with an increase in telomere resecting by MRN, due to the reduction in *SP100* expression, and an increase in expression of certain a-EJ genes, such as *LIG3* and *XRCC1*, may result in an increased frequency of a-NHEJ-mediate fusion of short sister-telomeres as an intermediate steps for activation of ALT. The resecting by a-EJ and recruitment of RPA1 may trigger the strand invasion between sister-telomeres or other telomeric sequences. Loss of functional ATRX may facilitate the strand invasion between sister-telomeres by suppressing the resolution of telomere cohesion. The upregulation of *BLM*, *RPA1,* and *BRCA2* may be required to process the telomere ends with one of HR pathways at the early post-crisis CPs.
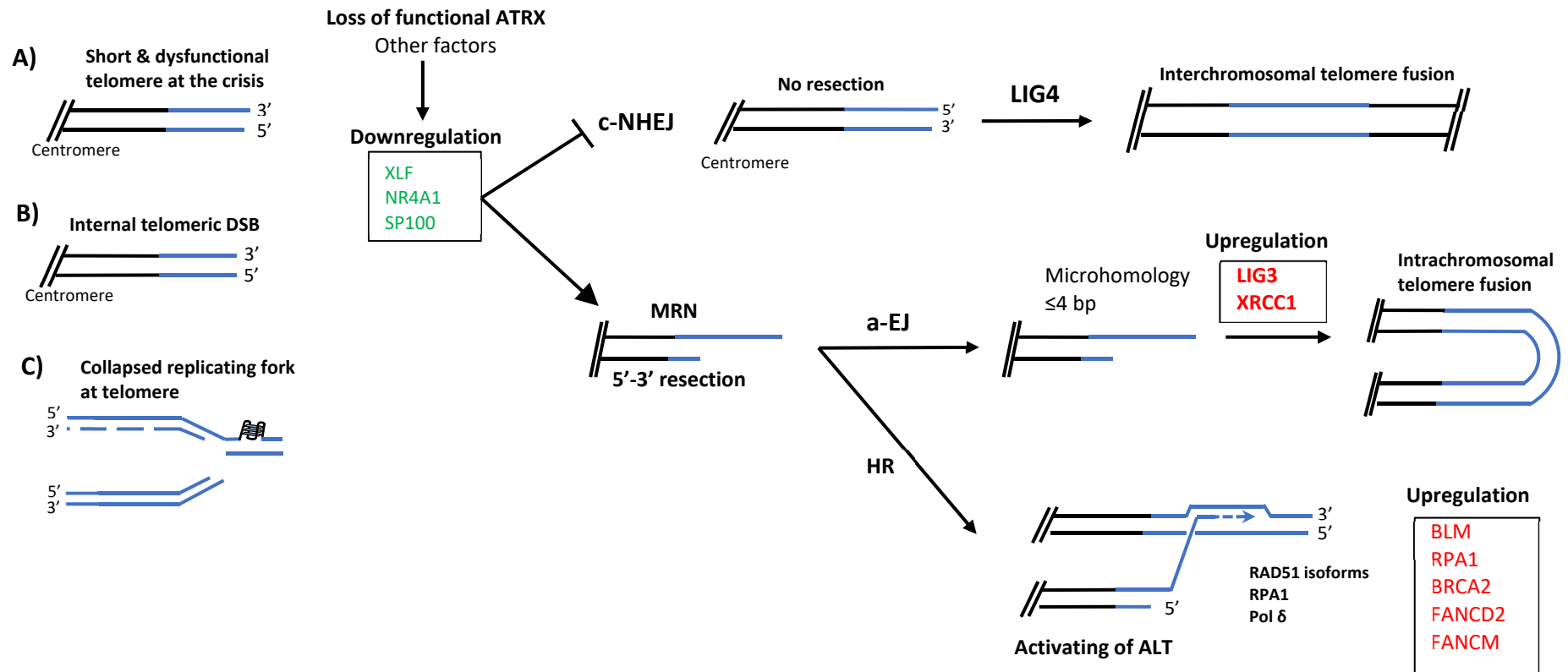
Figure 6.20 Model of ALT activation after ATRX knockout in pre-crisis SV40-transformed fibroblasts. The diagram illustrates the changes in gene expression that facilitate activation of ALT after ATRX knockout in SV40-tranformed fibroblasts. This model shows the processing of (A) the short and dysfunctional telomeres at the crisis, (B) the internal telomeric DSB that may result from break-fusion-bridge cycles during the crisis, (C) the collapsed replication fork at the lagging (G-rich) strand due to the replication stress that may result from loss of functional ATRX. Telomeric sequence coloured blue, downregulated genes coloured green, and upregulated genes coloured in red.

Chapter 7: Final discussion and future work

7.1 Roles of ORs in ALT

The NR2F/C classes of ORs showed to binds to ALT+ telomeres with high (TCAGGG)n repeats content, changes the telomeric chromatin status in a way that facilitates interaction between telomeres so promoting strand invasion and ALT activity in APBs (Conomos *et al.*, 2012; Conomos *et al.*, 2014). In this work, the frequency of telomere molecules with $(TCAGGG)_n$ repeats were measured at three telomeres XpYp, 12q, 17p using STELA and Southern blot hybridisation. This showed that some XpYp telomeres in 60% (3/5) ALT+ cell lines lacked potential binding site for these ORs. The exception was the WI38VA13_2RA cell line, with (TCAGGG)n repeats in 76%, 80%, 90% in XpYp, 12q, and 17p telomeres respectively. The investigation of protein expression by western blot analysis showed variation in the level of NR2F2, but not NR2C2 between the ALT+ cell lines. NR2F2 expression was significantly higher (~ 2-fold) in WV cell line compare other ALT+ cell lines.

Furthermore, the telomere localisation analyses showed that both NR2F2 and NR2C2 frequently colocalised with telomeres in the WI38VA13_2RA cells, 91% and 42% respectively. This was consistent with the high abundance of $(TCAGGG)_n$ in this cell line, and the frequency of localisation at telomeres reported in other studies (Dejardin and Kingston, 2009; Conomos *et al.*, 2012). In contrast, the high level of NR2F2 expression in the WV cell line did not result in preferential localisation at telomeres, as only ~ 5% of telomeric foci were positive for NR2F2. The telomeres that lack potential binding sites or NR2F2 or related ORs are unlikely to be able to act as donors or recipients for telomere lengthening via the proposed OR-ZNF817-NuRd mechanism (Conomos *et al.*, 2014), indicating that there must be more than one mechanism for bringing telomeres together prior to ALT. Another suggested mechanism to bring telomeres together is via the RAD51-HOP2-MND1 complex, which has been shown to drive the homology search and bring telomere together in U2OS cells (Cho *et al.*, 2014; Dilley *et al.*, 2016). It is possible that some ALT+ cells may use both mechanisms.

The regulatory role of NR2F2 was investigated by down regulation of protein expression using siRNA in WV, WI38VA13, and U2OS cell lines and RNA-seq analysis to investigate transcription changes. This resulted in an increase in the proportion of cells in G1 in WV and WI38VA13 with some similar transcription changes in biological process related to growth signals but no observed changes in expression of genes related to DNA repair mechanisms. By contrast, depletion of NR2F2 in U2OS cells resulted in significant changes in expression of genes involved in DNA replication and repair mechanisms, which may explain the cell cycle arrest in G1 and G2 in these cells. Other phenotypic changes observed in U2OS included a significant increase in the proportion of cells with APBs, which was consistent with an increase in the proportion of cells in G2/M. Therefore, it would be sensible to investigate the effect of NR2F2 downregulation in DNA damage response markers in U2OS. Further, investigating the interaction between the cell cycle checkpoints and replication fork progression could help to understand the mechanisms in which the cell cycle arrested upon NR2F2 downregulation in U2OS cells.

7.2 Genetic and transcriptional changes associated with activation of ALT

In this work, CRISPR-Cas9 was used to introduce inactivating mutations in exon 9 of *ATRX* in pre-crisis SV40 transformed pre-crisis fibroblast cells. In total 6 of 165 isolated clones gave rise to clonal populations that emerged from the crisis, and all were showed to be $ATRX^{-/-}$. Loss of functional ATRX seemed to facilitate the immortalisation in these cells. For example, the immortalisation frequency was 1.35 $X10^{-7}$ in cells transfected with CRISPR-TS1 and 2.75 $X10^{-6}$ in cells with co-transfected with both CRISPR-TS1 and CRISPR-TS2, whereas no immortal clones emerged from the cells transfect with the CRISPR empty vector. All the post-crisis CPs activated ALT despite the presence of wild type *TERT*. The investigation of *TERT* expression by RNA-seq showed no detectable expression any CPs, suggesting persistent epigenetic silencing of *TERT* in these cells. Analysis of telomerase activity using the TRAP assay could be used to confirm the expression status of *TERT*. Furthermore, it would be interesting to investigate the epigenetic markers such DNA methylation at the *TERT* promotor region.

In this work, genomic instability was investigated at early and late post-crisis CPs by measuring the frequency of micronuclei. The micronuclei analysis showed a high level of genomic instability associated with activation of ALT at an early post-crisis time point; 23-26% in CP_05, CP_15, CP_16, and 15-16% in CP_21, CP_38, CP_105. The estimation of ploidy showed average of three at early post-crisis compared to pre-crisis that demonstrate relatively stable genome and with estimated majority of cells are diploid cells. This was consistent with genomic instability associated with telomere-driven fusion events and tetraploidization during the crisis.

The genetic and transcriptional changes associated with activation of ALT were investigated in post-crisis CPs. Differential gene expression analysis at early post-crisis time points in emerging clonal populations showed that downregulation of the *JAK-STAT* signalling pathway, the ALT-suppressor *SP100* and reduced expression of c-NHEJ related genes *NR4A1* and *XLF* were the most remarkable events. The genetic changes associated with *SP100* downregulation seemed to be LOH at chr11 and chr19, while the reasons behind downregulation of *NR4A1* and *XLF* are still unclear. Further analysis of gene expression networks and epigenetic changes may be required to understand the mechanism in which the *NR4A1* and *XLF* genes have been downregulated.

ATRX is known to contribute to gene silencing by maintaining the heterochromatin status at the GC rich repetitive regions therefore, it would interesting to study the epigenetic markers at these regions. Finally, it would be useful to use ectopic expression to restore ATRX expression in post-crisis CPs cells to see which if any phenotype could be reversed.

References

Aaronson, D.S. and Horvath, C.M. (2002). A road map for those who don't know JAK-STAT. *Science (New York, N.Y.),* 296(5573), pp. 1653-1655.

Ahnesorg, P., Smith, P. and Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell,* 124(2), pp. 301-313.

Allshire, R.C., Dempster, M. and Hastie, N.D. (1989). Human telomeres contain at least three types of G-rich repeat distributed non-randomly. *Nucleic acids research,* 17(12), pp. 4611-4627.

Allsopp, R.C., Vaziri, H., Patterson, C., Goldstein, S., Younglai, E.V., Futcher, A.B., Greider, C.W. and Harley, C.B. (1992). Telomere length predicts replicative capacity of human fibroblasts. *Proceedings of the National Academy of Sciences of the United States of America,* 89(21), pp. 10114-10118.

Andres, S.N., Modesti, M., Tsai, C.J., Chu, G. and Junop, M.S. (2007). Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Molecular cell,* 28(6), pp. 1093-1101.

Annecke, K., Schmitt, M., Euler, U., Zerm, M., Paepke, D., Paepke, S., von Minckwitz, G., Thomssen, C. and Harbeck, N. (2008). uPA and PAI-1 in breast cancer: review of their clinical utility and current validation in the prospective NNBC-3 trial. *Advances in Clinical Chemistry,* 45, pp. 31-45.

Arora, R., Lee, Y., Wischnewski, H., Brun, C.M., Schwarz, T. and Azzalin, C.M. (2014). RNaseH1 regulates TERRA-telomeric DNA hybrids and telomere maintenance in ALT tumour cells. *Nature communications,* 5, pp. 5220.

Audebert, M., Salles, B. and Calsou, P. (2004). Involvement of poly(ADP-ribose) polymerase-1 and XRCC1/DNA ligase III in an alternative route for DNA double-strand breaks rejoining. *The Journal of biological chemistry,* 279(53), pp. 55117-55126.

Azzalin, C.M., Reichenbach, P., Khoriauli, L., Giulotto, E. and Lingner, J. (2007). Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (New York, N.Y.),* 318(5851), pp. 798-801.

Baird, D.M., Jeffreys, A.J. and Royle, N.J. (1995). Mechanisms underlying telomere repeat turnover, revealed by hypervariable variant repeat distribution patterns in the human Xp/Yp telomere. *The EMBO journal,* 14(21), pp. 5433-5443.

Baird, D.M., Rowson, J., Wynford-Thomas, D. and Kipling, D. (2003). Extensive allelic variation and ultrashort telomeres in senescent human cells. *Nature genetics,* 33(2), pp. 203-207.

Balk, B., Maicher, A., Dees, M., Klermund, J., Luke-Glaser, S., Bender, K. and Luke, B. (2013). Telomeric RNA-DNA hybrids affect telomere-length dynamics and senescence. *Nature structural & molecular biology,* 20(10), pp. 1199-1205.

Barber, G.N. (2000). The interferons and cell death: guardians of the cell or accomplices of apoptosis? *Seminars in cancer biology,* 10(2), pp. 103-111.

Bargonetti, J., Reynisdottir, I., Friedman, P.N. and Prives, C. (1992). Site-specific binding of wild-type p53 to cellular DNA is inhibited by SV40 T antigen and mutant p53. *Genes & development,* 6(10), pp. 1886-1898.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A. and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.),* 315(5819), pp. 1709-1712.

Baumann, P. and Cech, T.R. (2001). Pot1, the putative telomere end-binding protein in fission yeast and humans. *Science (New York, N.Y.),* 292(5519), pp. 1171-1175.

Bennardo, N., Cheng, A., Huang, N. and Stark, J.M. (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS genetics,* 4(6), pp. e1000110.

Benoit, G., Cooney, A., Giguere, V., Ingraham, H., Lazar, M., Muscat, G., Perlmann, T., Renaud, J.P., Schwabe, J., Sladek, F., Tsai, M.J. and Laudet, V. (2006). International Union of Pharmacology. LXVI. Orphan nuclear receptors. *Pharmacological reviews,* 58(4), pp. 798-836.

Bernstein, K.A. and Rothstein, R. (2009). At loose ends: resecting a double-strand break. *Cell,* 137(5), pp. 807-810.

Bhargava, R., Onyango, D.O. and Stark, J.M. (2016). Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends in genetics : TIG,* 32(9), pp. 566-575.

Bhargava, R., Sandhu, M., Muk, S., Lee, G., Vaidehi, N. and Stark, J.M. (2018). C-NHEJ without indels is robust and requires synergistic function of distinct XLF domains. *Nature communications,* 9(1), pp. 2484-018-04867-5.

Bhattacharyya, S., Keirsey, J., Russell, B., Kavecansky, J., Lillard-Wetherell, K., Tahmaseb, K., Turchi, J.J. and Groden, J. (2009). Telomerase-associated protein 1, HSP90, and topoisomerase IIalpha associate directly with the BLM helicase in immortalized cells using ALT and modulate its helicase activity using telomeric DNA substrates. *The Journal of biological chemistry,* 284(22), pp. 14966-14977.

Birmingham, A., Anderson, E., Sullivan, K., Reynolds, A., Boese, Q., Leake, D., Karpilow, J. and Khvorova, A. (2007). A protocol for designing siRNAs with high functionality and specificity. *Nature protocols,* 2(9), pp. 2068-2078.

Blackburn, E.H. and Gall, J.G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *Journal of Molecular Biology,* 120(1), pp. 33-53.

Boddy, M.N., Gaillard, P.H., McDonald, W.H., Shanahan, P., Yates, J.R.,3rd and Russell, P. (2001). Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell,* 107(4), pp. 537-548.

Borde, V. and Cobb, J. (2009). Double functions for the Mre11 complex during DNA double-strand break repair and replication. *The international journal of biochemistry & cell biology,* 41(6), pp. 1249-1253.

Borden, K.L. (2002). Pondering the promyelocytic leukemia protein (PML) puzzle: possible functions for PML nuclear bodies. *Molecular and cellular biology,* 22(15), pp. 5259-5269.

Bower, K., Napier, C.E., Cole, S.L., Dagg, R.A., Lau, L.M., Duncan, E.L., Moy, E.L. and Reddel, R.R. (2012). Loss of wild-type ATRX expression in somatic cell hybrids segregates with activation of Alternative Lengthening of Telomeres. *PloS one,* 7(11), pp. e50062.

Bridgham, J.T., Eick, G.N., Larroux, C., Deshpande, K., Harms, M.J., Gauthier, M.E., Ortlund, E.A., Degnan, B.M. and Thornton, J.W. (2010). Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS biology,* 8(10), pp. 10.1371/journal.pbio.1000497.

Broad Institute, G.R. (2016). Picard Toolkit.

Bryan, T.M., Englezou, A., Gupta, J., Bacchetti, S. and Reddel, R.R. (1995). Telomere elongation in immortal human cells without detectable telomerase activity. 14, pp. 4240-4248.

Bugreev, D.V., Huang, F., Mazina, O.M., Pezza, R.J., Voloshin, O.N., Camerini-Otero, R.D. and Mazin, A.V. (2014). HOP2-MND1 modulates RAD51 binding to nucleotides and DNA. *Nature communications,* 5, pp. 4198.

Cai, L., Yuan, W., Zhang, Z., He, L. and Chou, K.C. (2016). In-depth comparison of somatic point mutation callers based on different tumor next-generation sequencing depth data. *Scientific reports,* 6, pp. 36540.

Callebaut, I., Malivert, L., Fischer, A., Mornon, J.P., Revy, P. and de Villartay, J.P. (2006). Cernunnos interacts with the XRCC4 x DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor Nej1. *The Journal of biological chemistry,* 281(20), pp. 13857-13860.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R.S. and Zhang, Y. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science (New York, N.Y.), * 298(5595), pp. 1039-1043.

Cardoso, C., Timsit, S., Villard, L., Khrestchatisky, M., Fontes, M. and Colleaux, L. (1998). Specific interaction between the XNP/ATR-X gene product and the SET domain of the human EZH2 protein. *Human molecular genetics, * 7(4), pp. 679-684.

Carreira, A. and Kowalczykowski, S.C. (2011). Two classes of BRC repeats in BRCA2 promote RAD51 nucleoprotein filament function by distinct mechanisms. *Proceedings of the National Academy of Sciences of the United States of America, * 108(26), pp. 10448-10453.

Celli, G.B. and de Lange, T. (2005). DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. 7, pp. 712-718.

Cesare, A.J. and Griffith, J.D. (2004). Telomeric DNA in ALT cells is characterized by free telomeric circles and heterogeneous t-loops. *Molecular and cellular biology, * 24(22), pp. 9948-9957.

Cesare, A.J. and Reddel, R.R. (2010). Alternative lengthening of telomeres: models, mechanisms and implications. *Nature reviews.Genetics, * 11(5), pp. 319-330.

Chandra, V., Huang, P., Hamuro, Y., Raghuram, S., Wang, Y., Burris, T.P. and Rastinejad, F. (2008). Structure of the intact PPAR-gamma-RXR- nuclear receptor complex on DNA. *Nature, * 456(7220), pp. 350-356.

Chapman, J.R., Barral, P., Vannier, J.B., Borel, V., Steger, M., Tomas-Loba, A., Sartori, A.A., Adams, I.R., Batista, F.D. and Boulton, S.J. (2013). RIF1 is essential for 53BP1-dependent nonhomologous end joining and suppression of DNA double-strand break resection. *Molecular cell, * 49(5), pp. 858-871.

Chen, H.Z., Liu, Q.F., Li, L., Wang, W.J., Yao, L.M., Yang, M., Liu, B., Chen, W., Zhan, Y.Y., Zhang, M.Q., Cai, J.C., Zheng, Z.H., Lin, S.C., Li, B.A. and Wu, Q. (2012). The orphan receptor TR3 suppresses intestinal tumorigenesis in mice by downregulating Wnt signalling. *Gut, * 61(5), pp. 714-724.

Chen, L.Y., Redon, S. and Lingner, J. (2012). The human CST complex is a terminator of telomerase activity. *Nature, * 488(7412), pp. 540-544.

Chen, S.H., Plank, J.L., Willcox, S., Griffith, J.D. and Hsieh, T.S. (2014). Top3alpha is required during the convergent migration step of double Holliday junction dissolution. *PloS one, * 9(1), pp. e83582.

Chen, X., Qin, J., Cheng, C.M., Tsai, M.J. and Tsai, S.Y. (2012). COUP-TFII is a major regulator of cell cycle and Notch signaling pathways. *Molecular endocrinology (Baltimore, Md.), * 26(8), pp. 1268-1277.

Chen, X.B., Melchionna, R., Denis, C.M., Gaillard, P.H., Blasina, A., Van de Weyer, I., Boddy, M.N., Russell, P., Vialard, J. and McGowan, C.H. (2001). Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Molecular cell,* 8(5), pp. 1117-1127.

Cho, N.W., Dilley, R.L., Lampson, M.A. and Greenberg, R.A. (2014). Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell,* 159(1), pp. 108-121.

Chu, H.P., Cifuentes-Rojas, C., Kesner, B., Aeby, E., Lee, H.G., Wei, C., Oh, H.J., Boukhali, M., Haas, W. and Lee, J.T. (2017). TERRA RNA Antagonizes ATRX and Protects Telomeres. *Cell,* 170(1), pp. 86-101.e16.

Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S. and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature biotechnology,* 31(3), pp. 213-219.

Claessens, F. and Gewirth, D.T. (2004). DNA recognition by nuclear receptors. *Essays in biochemistry,* 40, pp. 59-72.

Clynes, D., Jelinska, C., Xella, B., Ayyub, H., Scott, C., Mitson, M., Taylor, S., Higgs, D.R. and Gibbons, R.J. (2015). Suppression of the alternative lengthening of telomere pathway by the chromatin remodelling factor ATRX. *Nature communications,* 6, pp. 7538.

Colgin, L.M. and Reddel, R.R. (1999). Telomere maintenance mechanisms and cellular immortalization. *Current opinion in genetics & development,* 9(1), pp. 97-103.

Cong, Y.S., Wen, J. and Bacchetti, S. (1999). The human telomerase catalytic subunit hTERT: organization of the gene and characterization of the promoter. *Human molecular genetics,* 8(1), pp. 137-142.

Conomos, D., Reddel, R.R. and Pickett, H.A. (2014). NuRD-ZNF827 recruitment to telomeres creates a molecular scaffold for homologous recombination. *Nature structural & molecular biology,* 21(9), pp. 760-770.

Conomos, D., Stutz, M.D., Hills, M., Neumann, A.A., Bryan, T.M., Reddel, R.R. and Pickett, H.A. (2012). Variant repeats are interspersed throughout the telomeres and recruit nuclear receptors in ALT cells. *The Journal of Cell Biology,* 199, pp. 893-906.

Cooney, A.J., Tsai, S.Y., O'Malley, B.W. and Tsai, M.J. (1992). Chicken ovalbumin upstream promoter transcription factor (COUP-TF) dimers bind to different GGTCA response elements, allowing COUP-TF to repress hormonal induction of the vitamin D3, thyroid hormone, and retinoic acid receptors. *Molecular and cellular biology,* 12(9), pp. 4153-4163.

Costantino, L., Sotiriou, S.K., Rantala, J.K., Magin, S., Mladenov, E., Helleday, T., Haber, J.E., Iliakis, G., Kallioniemi, O.P. and Halazonetis, T.D. (2014). Break-induced replication repair of damaged forks induces genomic duplications in human cells. *Science (New York, N.Y.),* 343(6166), pp. 88-91.

Counter, C.M., Avilion, A.A., LeFeuvre, C.E., Stewart, N.G., Greider, C.W., Harley, C.B. and Bacchetti, S. (1992). Telomere shortening associated with chromosome instability is arrested in immortal cells which express telomerase activity. *The EMBO journal,* 11(5), pp. 1921-1929.

Crasta, K., Ganem, N.J., Dagher, R., Lantermann, A.B., Ivanova, E.V., Pan, Y., Nezi, L., Protopopov, A., Chowdhury, D. and Pellman, D. (2012). DNA breaks and chromosome pulverization from errors in mitosis. *Nature,* 482(7383), pp. 53-58.

Cukusic, A., Skrobot Vidacek, N., Sopta, M. and Rubelj, I. (2008). Telomerase regulation at the crossroads of cell fate. *Cytogenetic and genome research,* 122(3-4), pp. 263-272.

Davoli, T. and de Lange, T. (2012). Telomere-driven tetraploidization occurs in human cells undergoing crisis and promotes transformation of mouse cells. *Cancer cell,* 21(6), pp. 765-776.

Davoli, T. and de Lange, T. (2011). The causes and consequences of polyploidy in normal development and cancer. *Annual Review of Cell and Developmental Biology,* 27, pp. 585-610.

Davoli, T., Denchi, E.L. and de Lange, T. (2010). Persistent telomere damage induces bypass of mitosis and tetraploidy. *Cell,* 141(1), pp. 81-93.

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

Deeb, K.K., Michalowska, A.M., Yoon, C.Y., Krummey, S.M., Hoenerhoff, M.J., Kavanaugh, C., Li, M.C., Demayo, F.J., Linnoila, I., Deng, C.X., Lee, E.Y., Medina, D., Shih, J.H. and Green, J.E. (2007). Identification of an integrated SV40 T/t-antigen cancer signature in aggressive human breast, prostate, and lung carcinomas with poor prognosis. *Cancer research,* 67(17), pp. 8065-8080.

Degasperi, A., Birtwistle, M.R., Volinsky, N., Rauch, J., Kolch, W. and Kholodenko, B.N. (2014). Evaluating strategies to normalise biological replicates of Western blot data. *PloS one,* 9(1), pp. e87293.

Dejardin, J. and Kingston, R.E. (2009). Purification of proteins associated with specific genomic Loci. *Cell,* 136(1), pp. 175-186.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J. and Charpentier, E. (2011). CRISPR RNA maturation by

trans-encoded small RNA and host factor RNase III. *Nature,* 471(7340), pp. 602-607.

Denchi, E.L. and de Lange, T. (2007). Protection of telomeres through independent control of ATM and ATR by TRF2 and POT1. 448, pp. 1068-1071.

Deng, S.K., Gibb, B., de Almeida, M.J., Greene, E.C. and Symington, L.S. (2014). RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nature structural & molecular biology,* 21(4), pp. 405-412.

Dilley, R.L., Verma, P., Cho, N.W., Winters, H.D., Wondisford, A.R. and Greenberg, R.A. (2016). Break-induced telomere synthesis underlies alternative telomere maintenance. *Nature,* 539(7627), pp. 54-58.

Dlaska, M., Anderl, C., Eisterer, W. and Bechter, O.E. (2008). Detection of circular telomeric DNA without 2D gel electrophoresis. *DNA and cell biology,* 27(9), pp. 489-496.

Donnianni, R.A. and Symington, L.S. (2013). Break-induced replication occurs by conservative DNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America,* 110(33), pp. 13475-13480.

Drane, P., Ouararhni, K., Depaux, A., Shuaib, M. and Hamiche, A. (2010). The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. 24, pp. 1253-1265.

Draskovic, I., Arnoult, N., Steiner, V., Bacchetti, S., Lomonte, P. and Londono-Vallejo, A. (2009). Probing PML body function in ALT cells reveals spatiotemporal requirements for telomere recombination. *Proceedings of the National Academy of Sciences of the United States of America,* 106(37), pp. 15726-15731.

Dunham, M.A., Neumann, A.A., Fasching, C.L. and Reddel, R.R. (2000). Telomere maintenance by recombination in human cells. 26, pp. 447-450.

El Ghouzzi, V., Dagoneau, N., Kinning, E., Thauvin-Robinet, C., Chemaitilly, W., Prost-Squarcioni, C., Al-Gazali, L.I., Verloes, A., Le Merrer, M., Munnich, A., Trembath, R.C. and Cormier-Daire, V. (2003). Mutations in a novel gene Dymeclin (FLJ20071) are responsible for Dyggve-Melchior-Clausen syndrome. *Human molecular genetics,* 12(3), pp. 357-364.

Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K. and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature,* 411(6836), pp. 494-498.

Elbashir, S.M., Lendeckel, W. and Tuschl, T. (2001). RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & development,* 15(2), pp. 188-200.

Elsasser, S.J., Noh, K.M., Diaz, N., Allis, C.D. and Banaszynski, L.A. (2015). Histone H3.3 is required for endogenous retroviral element silencing in embryonic stem cells. *Nature,* 522(7555), pp. 240-244.

Episkopou, H., Draskovic, I., Van Beneden, A., Tilman, G., Mattiussi, M., Gobin, M., Arnoult, N., Londono-Vallejo, A. and Decottignies, A. (2014). Alternative Lengthening of Telomeres is characterized by reduced compaction of telomeric chromatin. *Nucleic acids research,* 42(7), pp. 4391-4405.

Eustermann, S., Yang, J.C., Law, M.J., Amos, R., Chapman, L.M., Jelinska, C., Garrick, D., Clynes, D., Gibbons, R.J., Rhodes, D., Higgs, D.R. and Neuhaus, D. (2011). Combinatorial readout of histone H3 modifications specifies localization of ATRX to heterochromatin. *Nature structural & molecular biology,* 18(7), pp. 777-782.

Fan, Q., Zhang, F., Barrett, B., Ren, K. and Andreassen, P.R. (2009). A role for monoubiquitinated FANCD2 at telomeres in ALT cells. *Nucleic acids research,* 37(6), pp. 1740-1754.

Fasching, C.L., Neumann, A.A., Muntoni, A., Yeager, T.R. and Reddel, R.R. (2007). DNA damage induces alternative lengthening of telomeres (ALT) associated promyelocytic leukemia bodies that preferentially associate with linear telomeric DNA. *Cancer research,* 67(15), pp. 7072-7077.

Favero, F., Joshi, T., Marquard, A.M., Birkbak, N.J., Krzystanek, M., Li, Q., Szallasi, Z. and Eklund, A.C. (2015). Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Annals of oncology : official journal of the European Society for Medical Oncology,* 26(1), pp. 64-70.

Feinberg, A.P. and Vogelstein, B. (1984). "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Addendum. *Analytical Biochemistry,* 137(1), pp. 266-267.

Feng, X., Hsu, S.J., Bhattacharjee, A., Wang, Y., Diao, J. and Price, C.M. (2018). CTC1-STN1 terminates telomerase while STN1-TEN1 enables C-strand synthesis during telomere replication in colon cancer cells. *Nature communications,* 9(1), pp. 2827-018-05154-z.

Filion, C., Motoi, T., Olshen, A.B., Lae, M., Emnett, R.J., Gutmann, D.H., Perry, A., Ladanyi, M. and Labelle, Y. (2009). The EWSR1/NR4A3 fusion protein of extraskeletal myxoid chondrosarcoma activates the PPARG nuclear receptor gene. *The Journal of pathology,* 217(1), pp. 83-93.

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. and Zou, L. (2015). Alternative lengthening of telomeres renders cancer cells hypersensitive to ATR inhibitors. *Science (New York, N.Y.),* 347(6219), pp. 273-277.

Fouquin, A., Guirouilh-Barbat, J., Lopez, B., Hall, J., Amor-Gueret, M. and Pennaneach, V. (2017). PARP2 controls double-strand break repair pathway choice by limiting 53BP1 accumulation at DNA damage sites and promoting end-resection. *Nucleic acids research,* 45(21), pp. 12325-12339.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H. and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature,* 468(7320), pp. 67-71.

Garrick, D., Samara, V., McDowell, T.L., Smith, A.J., Dobbie, L., Higgs, D.R. and Gibbons, R.J. (2004). A conserved truncated isoform of the ATR-X syndrome protein lacking the SWI/SNF-homology domain. *Gene,* 326, pp. 23-34.

Gibbons, R.J., Suthers, G.K., Wilkie, A.O., Buckle, V.J. and Higgs, D.R. (1992). X-linked alpha-thalassemia/mental retardation (ATR-X) syndrome: localization to Xq12-q21.31 by X inactivation and linkage analysis. *American Journal of Human Genetics,* 51(5), pp. 1136-1149.

Giguere, V. (1999). Orphan nuclear receptors: from gene to function. *Endocrine reviews,* 20(5), pp. 689-725.

Goldberg, A.D., Banaszynski, L.A., Noh, K.M., Lewis, P.W., Elsaesser, S.J., Stadler, S., Dewell, S., Law, M., Guo, X., Li, X., Wen, D., Chapgier, A., DeKelver, R.C., Miller, J.C., Lee, Y.L., Boydston, E.A., Holmes, M.C., Gregory, P.D., Greally, J.M., Rafii, S., Yang, C., Scambler, P.J., Garrick, D., Gibbons, R.J., Higgs, D.R., Cristea, I.M., Urnov, F.D., Zheng, D. and Allis, C.D. (2010). Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell,* 140(5), pp. 678-691.

Goldstein, S. (1990). Replicative senescence: the human fibroblast comes of age. *Science (New York, N.Y.),* 249(4973), pp. 1129-1133.

Gongora, C., David, G., Pintard, L., Tissot, C., Hua, T.D., Dejean, A. and Mechti, N. (1997). Molecular cloning of a new interferon-induced PML nuclear body-associated protein. *The Journal of biological chemistry,* 272(31), pp. 19457-19463.

Goto, M., Rubenstein, M., Weber, J., Woods, K. and Drayna, D. (1992). Genetic linkage of Werner's syndrome to five markers on chromosome 8. *Nature,* 355(6362), pp. 735-738.

Greider, C.W. (1996). Telomere length regulation. *Annual Review of Biochemistry,* 65, pp. 337-365.

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H. and de Lange, T. (1999). Mammalian telomeres end in a large duplex loop. *Cell,* 97, pp. 503-514.

Grobelny, J.V., Godwin, A.K. and Broccoli, D. (2000). ALT-associated PML bodies are present in viable cells and are enriched in cells in the G(2)/M phase of the cell cycle. *Journal of cell science,* 113 Pt 24, pp. 4577-4585.

Grotzinger, T., Jensen, K. and Will, H. (1996). The interferon (IFN)-stimulated gene Sp100 promoter contains an IFN-gamma activation site and an imperfect IFN-stimulated response element which mediate type I IFN inducibility. *The Journal of biological chemistry,* 271(41), pp. 25253-25260.

Grudic, A., Jul-Larsen, A., Haring, S.J., Wold, M.S., Lonning, P.E., Bjerkvig, R. and Boe, S.O. (2007). Replication protein A prevents accumulation of single-stranded telomeric DNA in cells that use alternative lengthening of telomeres. *Nucleic acids research,* 35(21), pp. 7267-7278.

Gunduz, V., Kong, E., Bryan, C.D. and Hinds, P.W. (2012). Loss of the retinoblastoma tumor suppressor protein in murine calvaria facilitates immortalization of osteoblast-adipocyte bipotent progenitor cells characterized by low expression of N-cadherin. *Molecular and cellular biology,* 32(13), pp. 2561-2569.

Haft, D.H., Selengut, J., Mongodin, E.F. and Nelson, K.E. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS computational biology,* 1(6), pp. e60.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M. and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. *Cell,* 139(5), pp. 945-956.

Hammond, S.M., Bernstein, E., Beach, D. and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature,* 404(6775), pp. 293-296.

Harding, H.P. and Lazar, M.A. (1993). The orphan receptor Rev-ErbA alpha activates transcription via a novel response element. *Molecular and cellular biology,* 13(5), pp. 3113-3121.

HAYFLICK, L. (1965). The Limited in Vitro Lifetime of Human Diploid Cell Strains. *Experimental cell research,* 37, pp. 614-636.

Heaphy, C.M., de Wilde, R.F., Jiao, Y., Klein, A.P., Edil, B.H., Shi, C., Bettegowda, C., Rodriguez, F.J., Eberhart, C.G., Hebbar, S., Offerhaus, G.J., McLendon, R., Rasheed, B.A., He, Y., Yan, H., Bigner, D.D., Oba-Shinjo, S.M., Marie, S.K., Riggins, G.J., Kinzler, K.W., Vogelstein, B., Hruban, R.H., Maitra, A., Papadopoulos, N. and Meeker, A.K. (2011). Altered telomeres in tumors with ATRX and DAXX mutations. *Science (New York, N.Y.),* 333(6041), pp. 425.

Heaphy, C.M., Subhawong, A.P., Hong, S.M., Goggins, M.G., Montgomery, E.A., Gabrielson, E., Netto, G.J., Epstein, J.I., Lotan, T.L., Westra, W.H., Shih, I., Iacobuzio-Donahue, C.A., Maitra, A., Li, Q.K., Eberhart, C.G., Taube, J.M.,

Rakheja, D., Kurman, R.J., Wu, T.C., Roden, R.B., Argani, P., De Marzo, A.M., Terracciano, L., Torbenson, M. and Meeker, A.K. (2011). Prevalence of the alternative lengthening of telomeres telomere maintenance mechanism in human cancer subtypes. *The American journal of pathology,* 179(4), pp. 1608-1615.

Henson, J.D., Cao, Y., Huschtscha, L.I., Chang, A.C., Au, A.Y., Pickett, H.A. and Reddel, R.R. (2009). DNA C-circles are specific and quantifiable markers of alternative-lengthening-of-telomeres activity. 27, pp. 1181-1185.

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. and 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. 11, pp. 217-225.

Henson, J.D., Neumann, A.A., Yeager, T.R. and Reddel, R.R. (2002). Alternative lengthening of telomeres in mammalian cells. 21, pp. 598-610.

Henson, J.D. and Reddel, R.R. (2010). Assaying and investigating Alternative Lengthening of Telomeres activity in human cells and cancers. *FEBS letters,* 584(17), pp. 3800-3811.

Hernandez, P. and Tirnauer, J.S. (2010). Tumor suppressor interactions with microtubules: keeping cell polarity and cell division on track. *Disease models & mechanisms,* 3(5-6), pp. 304-315.

Hom, R.A. and Wuttke, D.S. (2017). Human CST Prefers G-Rich but Not Necessarily Telomeric Sequences. *Biochemistry,* 56(32), pp. 4210-4218.

Huang da, W., Sherman, B.T. and Lempicki, R.A. (2009). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research,* 37(1), pp. 1-13.

Huang, C., Jia, P., Chastain, M., Shiva, O. and Chai, W. (2017). The human CTC1/STN1/TEN1 complex regulates telomere maintenance in ALT cancer cells. *Experimental cell research,* 355(2), pp. 95-104.

Hughes, T.R., Weilbaecher, R.G., Walterscheid, M. and Lundblad, V. (2000). Identification of the single-strand telomeric DNA binding domain of the Saccharomyces cerevisiae Cdc13 protein. *Proceedings of the National Academy of Sciences of the United States of America,* 97(12), pp. 6457-6462.

Hung, Y.F., Chen, C.Y., Shih, Y.C., Liu, H.Y., Huang, C.M. and Hsueh, Y.P. (2018). Endosomal TLR3, TLR7, and TLR8 control neuronal morphology through different transcriptional programs. *The Journal of cell biology,* 217(8), pp. 2727-2742.

Iwase, S., Xiang, B., Ghosh, S., Ren, T., Lewis, P.W., Cochrane, J.C., Allis, C.D., Picketts, D.J., Patel, D.J., Li, H. and Shi, Y. (2011). ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nature structural & molecular biology,* 18(7), pp. 769-776.

Jansen, R., Embden, J.D., Gaastra, W. and Schouls, L.M. (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Molecular microbiology,* 43(6), pp. 1565-1575.

Jatiani, S.S., Baker, S.J., Silverman, L.R. and Reddy, E.P. (2010). Jak/STAT pathways in cytokine signaling and myeloproliferative disorders: approaches for targeted therapies. *Genes & cancer,* 1(10), pp. 979-993.

Jaworski, C.J., Moreira, E., Li, A., Lee, R. and Rodriguez, I.R. (2001). A family of 12 human genes containing oxysterol-binding domains. *Genomics,* 78(3), pp. 185-196.

Jazayeri, A., Falck, J., Lukas, C., Bartek, J., Smith, G.C., Lukas, J. and Jackson, S.P. (2006). ATM- and cell cycle-dependent regulation of ATR in response to DNA double-strand breaks. *Nature cell biology,* 8(1), pp. 37-45.

Jiang, H., Lei, R., Ding, S.W. and Zhu, S. (2014). Skewer: a fast and accurate adapter trimmer for next-generation sequencing paired-end reads. *BMC bioinformatics,* 15, pp. 182-2105-15-182.

Jiang, L., Kon, N., Li, T., Wang, S.J., Su, T., Hibshoosh, H., Baer, R. and Gu, W. (2015). Ferroptosis as a p53-mediated activity during tumour suppression. *Nature,* 520(7545), pp. 57-62.

Jiang, W.Q., Zhong, Z.H., Henson, J.D., Neumann, A.A., Chang, A.C. and Reddel, R.R. (2005). Suppression of alternative lengthening of telomeres by Sp100-mediated sequestration of the MRE11/RAD50/NBS1 complex. *Molecular and cellular biology,* 25(7), pp. 2708-2721.

Jiang, W.Q., Zhong, Z.H., Henson, J.D. and Reddel, R.R. (2007). Identification of candidate alternative lengthening of telomeres genes by methionine restriction and RNA interference. *Oncogene,* 26(32), pp. 4635-4647.

Jiao, Y., Killela, P.J., Reitman, Z.J., Rasheed, A.B., Heaphy, C.M., de Wilde, R.F., Rodriguez, F.J., Rosemberg, S., Oba-Shinjo, S.M., Nagahashi Marie, S.K., Bettegowda, C., Agrawal, N., Lipp, E., Pirozzi, C., Lopez, G., He, Y., Friedman, H., Friedman, A.H., Riggins, G.J., Holdhoff, M., Burger, P., McLendon, R., Bigner, D.D., Vogelstein, B., Meeker, A.K., Kinzler, K.W., Papadopoulos, N., Diaz, L.A. and Yan, H. (2012). Frequent ATRX, CIC, FUBP1 and IDH1 mutations refine the classification of malignant gliomas. *Oncotarget,* 3(7), pp. 709-722.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.),* 337(6096), pp. 816-821.

Johnson, D.G., Schwarz, J.K., Cress, W.D. and Nevins, J.R. (1993). Expression of transcription factor E2F1 induces quiescent cells to enter S phase. *Nature,* 365(6444), pp. 349-352.

Jones, R.E., Oh, S., Grimstead, J.W., Zimbric, J., Roger, L., Heppel, N.H., Ashelford, K.E., Liddiard, K., Hendrickson, E.A. and Baird, D.M. (2014). Escape from telomere-driven crisis is DNA ligase III dependent. *Cell reports,* 8(4), pp. 1063-1076.

Karlseder, J., Broccoli, D., Dai, Y., Hardy, S. and de Lange, T. (1999). p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. 283, pp. 1321-1325.

Kegel, A. and Sjogren, C. (2010). The Smc5/6 complex: more than repair? *Cold Spring Harbor symposia on quantitative biology,* 75, pp. 179-187.

Kim, S., Scheffler, K., Halpern, A.L., Bekritsky, M.A., Noh, E., Kallberg, M., Chen, X., Kim, Y., Beyter, D., Krusche, P. and Saunders, C.T. (2018). Strelka2: fast and accurate calling of germline and somatic variants. *Nature methods,* 15(8), pp. 591-594.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z. and Joung, J.K. (2016). High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature,* 529(7587), pp. 490-495.

Klobutcher, L.A., Swanton, M.T., Donini, P. and Prescott, D.M. (1981). All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3' terminus. *Proceedings of the National Academy of Sciences of the United States of America,* 78(5), pp. 3015-3019.

Kramer, K.M., Brock, J.A., Bloom, K., Moore, J.K. and Haber, J.E. (1994). Two different types of double-strand breaks in Saccharomyces cerevisiae are repaired by similar RAD52-independent, nonhomologous recombination events. *Molecular and cellular biology,* 14(2), pp. 1293-1301.

Kroigard, A.B., Thomassen, M., Laenkholm, A.V., Kruse, T.A. and Larsen, M.J. (2016). Evaluation of Nine Somatic Variant Callers for Detection of Somatic Mutations in Exome and Targeted Deep Sequencing Data. *PloS one,* 11(3), pp. e0151664.

Kyo, S., Takakura, M., Fujiwara, T. and Inoue, M. (2008). Understanding and exploiting hTERT promoter regulation for diagnosis and treatment of human cancers. *Cancer science,* 99(8), pp. 1528-1538.

Lang, M., Jegou, T., Chung, I., Richter, K., Munch, S., Udvarhelyi, A., Cremer, C., Hemmerich, P., Engelhardt, J., Hell, S.W. and Rippe, K. (2010). Three-dimensional organization of promyelocytic leukemia nuclear bodies. *Journal of cell science,* 123(Pt 3), pp. 392-400.

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. and Ding, L. (2012). SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics (Oxford, England),* 28(3), pp. 311-317.

Law, M.J., Lower, K.M., Voon, H.P., Hughes, J.R., Garrick, D., Viprakasit, V., Mitson, M., De Gobbi, M., Marra, M., Morris, A., Abbott, A., Wilder, S.P., Taylor, S., Santos, G.M., Cross, J., Ayyub, H., Jones, S., Ragoussis, J., Rhodes, D., Dunham, I., Higgs, D.R. and Gibbons, R.J. (2010). ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell,* 143(3), pp. 367-378.

Lechner, M.S., Schultz, D.C., Negorev, D., Maul, G.G. and Rauscher, F.J.,3rd (2005). The mammalian heterochromatin protein 1 binds diverse nuclear proteins through a common motif that targets the chromoshadow domain. *Biochemical and biophysical research communications,* 331(4), pp. 929-937.

Lee, M., Hills, M., Conomos, D., Stutz, M.D., Dagg, R.A., Lau, L.M., Reddel, R.R. and Pickett, H.A. (2014). Telomere extension by telomerase and ALT generates variant repeats by mechanistically distinct processes. *Nucleic acids research,* 42(3), pp. 1733-1746.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S. and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature,* 425(6956), pp. 415-419.

Lendvay, T.S., Morris, D.K., Sah, J., Balasubramanian, B. and Lundblad, V. (1996). Senescence mutants of Saccharomyces cerevisiae with a defect in telomere replication identify three additional EST genes. *Genetics,* 144(4), pp. 1399-1412.

Levy, M.A., Kernohan, K.D., Jiang, Y. and Berube, N.G. (2015). ATRX promotes gene expression by facilitating transcriptional elongation through guanine-rich coding regions. *Human molecular genetics,* 24(7), pp. 1824-1835.

Lewis, P.W., Elsaesser, S.J., Noh, K.M., Stadler, S.C. and Allis, C.D. (2010). Daxx is an H3.3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. *Proceedings of the National Academy of Sciences of the United States of America,* 107(32), pp. 14075-14080.

Li, B., Oestreich, S. and de Lange, T. (2000). Identification of human Rap1: implications for telomere evolution. *Cell,* 101(5), pp. 471-483.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England),* 25(14), pp. 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009). The

Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England),* 25(16), pp. 2078-2079.

Liao, Y., Smyth, G.K. and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic acids research,* 41(10), pp. e108.

Lieber, M.R. (2010). The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual Review of Biochemistry,* 79, pp. 181-211.

Lillard-Wetherell, K., Machwe, A., Langland, G.T., Combs, K.A., Behbehani, G.K., Schonberg, S.A., German, J., Turchi, J.J., Orren, D.K. and Groden, J. (2004). Association and regulation of the BLM helicase by the telomere proteins TRF1 and TRF2. *Human molecular genetics,* 13(17), pp. 1919-1932.

Lin, W., Sampathi, S., Dai, H., Liu, C., Zhou, M., Hu, J., Huang, Q., Campbell, J., Shin-Ya, K., Zheng, L., Chai, W. and Shen, B. (2013). Mammalian DNA2 helicase/nuclease cleaves G-quadruplex DNA and is required for telomere integrity. *The EMBO journal,* 32(10), pp. 1425-1439.

Linger, B.R. and Price, C.M. (2009). Conservation of telomere protein complexes: shuffling through evolution. *Critical reviews in biochemistry and molecular biology,* 44(6), pp. 434-446.

Litchfield, L.M. and Klinge, C.M. (2012). Multiple roles of COUP-TFII in cancer initiation and progression. *Journal of Molecular Endocrinology,* 49(3), pp. R135-48.

Love, M.I., Huber, W. and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology,* 15(12), pp. 550-014-0550-8.

Lovejoy, C.A., Li, W., Reisenweber, S., Thongthip, S., Bruno, J., de Lange, T., De, S., Petrini, J.H., Sung, P.A., Jasin, M., Rosenbluh, J., Zwang, Y., Weir, B.A., Hatton, C., Ivanova, E., Macconaill, L., Hanna, M., Hahn, W.C., Lue, N.F., Reddel, R.R., Jiao, Y., Kinzler, K., Vogelstein, B., Papadopoulos, N. and Meeker, A.K. (2012). Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. 8, pp. e1002772.

Luciani, J.J., Depetris, D., Usson, Y., Metzler-Guillemain, C., Mignon-Ravix, C., Mitchell, M.J., Megarbane, A., Sarda, P., Sirma, H., Moncla, A., Feunteun, J. and Mattei, M.G. (2006). PML nuclear bodies are highly organised DNA-protein structures with a function in heterochromatin remodelling at the G2 phase. *Journal of cell science,* 119(Pt 12), pp. 2518-2531.

Lundblad, V. and Blackburn, E.H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell,* 73(2), pp. 347-360.

Mabuchi, M., Kataoka, H., Miura, Y., Kim, T.S., Kawaguchi, M., Ebi, M., Tanaka, M., Mori, Y., Kubota, E., Mizushima, T., Shimura, T., Mizoshita, T., Tanida, S., Kamiya, T., Asai, K. and Joh, T. (2010). Tumor suppressor, AT motif binding factor 1 (ATBF1), translocates to the nucleus with runt domain transcription factor 3 (RUNX3) in response to TGF-beta signal transduction. *Biochemical and biophysical research communications,* 398(2), pp. 321-325.

Makharashvili, N., Tubbs, A.T., Yang, S.H., Wang, H., Barton, O., Zhou, Y., Deshpande, R.A., Lee, J.H., Lobrich, M., Sleckman, B.P., Wu, X. and Paull, T.T. (2014). Catalytic and noncatalytic roles of the CtIP endonuclease in double-strand break end resection. *Molecular cell,* 54(6), pp. 1022-1033.

Malewicz, M., Kadkhodaei, B., Kee, N., Volakakis, N., Hellman, U., Viktorsson, K., Leung, C.Y., Chen, B., Lewensohn, R., van Gent, D.C., Chen, D.J. and Perlmann, T. (2011). Essential role for DNA-PK-mediated phosphorylation of NR4A nuclear orphan receptors in DNA double-strand break repair. *Genes & development,* 25(19), pp. 2031-2040.

Mangelsdorf, D.J., Thummel, C., Beato, M., Herrlich, P., Schutz, G., Umesono, K., Blumberg, B., Kastner, P., Mark, M., Chambon, P. and Evans, R.M. (1995). The nuclear receptor superfamily: the second decade. *Cell,* 83(6), pp. 835-839.

Marciano, D.P., Chang, M.R., Corzo, C.A., Goswami, D., Lam, V.Q., Pascal, B.D. and Griffin, P.R. (2014). The therapeutic potential of nuclear receptor modulators for treatment of metabolic disorders: PPARgamma, RORs, and Rev-erbs. *Cell metabolism,* 19(2), pp. 193-208.

Marians, K.J. (2000). PriA-directed replication fork restart in Escherichia coli. *Trends in biochemical sciences,* 25(4), pp. 185-189.

Marraffini, L.A. and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science (New York, N.Y.),* 322(5909), pp. 1843-1845.

Martin, L.J. and Tremblay, J.J. (2010). Nuclear receptors in Leydig cell gene expression and function. *Biology of reproduction,* 83(1), pp. 3-14.

Marzec, P., Armenise, C., Perot, G., Roumelioti, F.M., Basyuk, E., Gagos, S., Chibon, F. and Dejardin, J. (2015). Nuclear-receptor-mediated telomere insertion leads to genome instability in ALT cancers. *Cell,* 160(5), pp. 913-927.

Maser, R.S., Wong, K.K., Sahin, E., Xia, H., Naylor, M., Hedberg, H.M., Artandi, S.E. and DePinho, R.A. (2007). DNA-dependent protein kinase catalytic subunit is not required for dysfunctional telomere fusion and checkpoint response in the telomerase-deficient mouse. *Molecular and cellular biology,* 27(6), pp. 2253-2265.

Mashal, R.D., Koontz, J. and Sklar, J. (1995). Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nature genetics,* 9(2), pp. 177-183.

Matos, J. and West, S.C. (2014). Holliday junction resolution: regulation in space and time. *DNA repair,* 19, pp. 176-181.

Mayakonda, A. & Koeffler, H.P. (2016). Maftools: Efficient analysis, visualization and summarization of MAF files from large-scale cohort based cancer studies. *bioRxiv, .*

McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics,* 26(2), pp. 234-282.

McDowell, T.L., Gibbons, R.J., Sutherland, H., O'Rourke, D.M., Bickmore, W.A., Pombo, A., Turley, H., Gatter, K., Picketts, D.J., Buckle, V.J., Chapman, L., Rhodes, D. and Higgs, D.R. (1999). Localization of a putative transcriptional regulator (ATRX) at pericentromeric heterochromatin and the short arms of acrocentric chromosomes. *Proceedings of the National Academy of Sciences of the United States of America,* 96(24), pp. 13983-13988.

McEachern, M.J. and Haber, J.E. (2006). Break-induced replication and recombinational telomere elongation in yeast. *Annual Review of Biochemistry,* 75, pp. 111-135.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. and DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research,* 20(9), pp. 1297-1303.

Meek, K., Dang, V. and Lees-Miller, S.P. (2008). DNA-PK: the means to justify the ends? *Advances in Immunology,* 99, pp. 33-58.

Mendez-Bermudez, A., Hidalgo-Bravo, A., Cotton, V.E., Gravani, A., Jeyapalan, J.N. and Royle, N.J. (2012). The roles of WRN and BLM RecQ helicases in the Alternative Lengthening of Telomeres. *Nucleic acids research,* 40(21), pp. 10809-10820.

Meyne, J., Ratliff, R.L. and Moyzis, R.K. (1989). Conservation of the human telomere sequence (TTAGGG)n among vertebrates. *Proceedings of the National Academy of Sciences of the United States of America,* 86(18), pp. 7049-7053.

Min, J., Wright, W.E. and Shay, J.W. (2017). Alternative lengthening of telomeres can be maintained by preferential elongation of lagging strands. *Nucleic acids research,* 45(5), pp. 2615-2628.

Miura, T., Yamana, Y., Usui, T., Ogawa, H.I., Yamamoto, M.T. and Kusano, K. (2012). Homologous recombination via synthesis-dependent strand annealing in yeast requires the Irc20 and Srs2 DNA helicases. *Genetics,* 191(1), pp. 65-78.

Miyake, Y., Nakamura, M., Nabetani, A., Shimamura, S., Tamura, M., Yonehara, S., Saito, M. and Ishikawa, F. (2009). RPA-like mammalian Ctc1-Stn1-Ten1 complex binds to single-stranded DNA and protects telomeres independently of the Pot1 pathway. *Molecular cell,* 36(2), pp. 193-206.

Moh, M.C. and Shen, S. (2009). The roles of cell adhesion molecules in tumor suppression and cell migration: a new paradox. *Cell adhesion & migration,* 3(4), pp. 334-336.

Mohaghegh, P., Karow, J.K., Brosh, R.M.,Jr, Bohr, V.A. and Hickson, I.D. (2001). The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic acids research,* 29(13), pp. 2843-2849.

Mohan, H.M., Aherne, C.M., Rogers, A.C., Baird, A.W., Winter, D.C. and Murphy, E.P. (2012). Molecular pathways: the role of NR4A orphan nuclear receptors in cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research,* 18(12), pp. 3223-3228.

Mojica, F.J., Diez-Villasenor, C., Soria, E. and Juez, G. (2000). Biological significance of a family of regularly spaced repeats in the genomes of Archaea, Bacteria and mitochondria. *Molecular microbiology,* 36(1), pp. 244-246.

Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M. and Valen, E. (2014). CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic acids research,* 42(Web Server issue), pp. W401-7.

Moon, A.F., Pryor, J.M., Ramsden, D.A., Kunkel, T.A., Bebenek, K. and Pedersen, L.C. (2014). Sustained active site rigidity during synthesis by human DNA polymerase mu. *Nature structural & molecular biology,* 21(3), pp. 253-260.

Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988). A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proceedings of the National Academy of Sciences of the United States of America,* 85(18), pp. 6622-6626.

Muller, H.J. (1938). The remaking of chromosomes. *The Collecting Net-Woods Hole,* 8(13), pp. 181-198.

Murnane, J.P., Sabatier, L., Marder, B.A. and Morgan, W.F. (1994). Telomere dynamics in an immortal human cell line. *The EMBO journal,* 13(20), pp. 4953-4962.

Nabetani, A., Yokoyama, O. and Ishikawa, F. (2004). Localization of hRad9, hHus1, hRad1, and hRad17 and caffeine-sensitive DNA replication at the alternative lengthening of telomeres-associated promyelocytic leukemia body. *The Journal of biological chemistry,* 279(24), pp. 25849-25857.

Nan, X., Hou, J., Maclean, A., Nasir, J., Lafuente, M.J., Shu, X., Kriaucionis, S. and Bird, A. (2007). Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proceedings of the National Academy of Sciences of the United States of America,* 104(8), pp. 2709-2714.

Napier, C.E., Huschtscha, L.I., Harvey, A., Bower, K., Noble, J.R., Hendrickson, E.A. and Reddel, R.R. (2015). ATRX represses alternative lengthening of telomeres. *Oncotarget, .*

Navab, R., Gonzalez-Santos, J.M., Johnston, M.R., Liu, J., Brodt, P., Tsao, M.S. and Hu, J. (2004). Expression of chicken ovalbumin upstream promoter-transcription factor II enhances invasiveness of human lung carcinoma cells. *Cancer research,* 64(15), pp. 5097-5105.

Navarro, P., Gomez, M., Pizarro, A., Gamallo, C., Quintanilla, M. and Cano, A. (1991). A role for the E-cadherin cell-cell adhesion molecule during tumor progression of mouse epidermal carcinogenesis. *The Journal of cell biology,* 115(2), pp. 517-533.

Nogales, E. (2000). Structural insights into microtubule function. *Annual Review of Biochemistry,* 69, pp. 277-302.

Nugent, C.I., Hughes, T.R., Lue, N.F. and Lundblad, V. (1996). Cdc13p: a single-strand telomeric DNA-binding protein with a dual role in yeast telomere maintenance. *Science (New York, N.Y.),* 274(5285), pp. 249-252.

Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. and Sugino, A. (1968). Mechanism of DNA chain growth. I. Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America,* 59(2), pp. 598-605.

Opresko, P.L., Otterlei, M., Graakjaer, J., Bruheim, P., Dawut, L., Kolvraa, S., May, A., Seidman, M.M. and Bohr, V.A. (2004). The Werner syndrome helicase and exonuclease cooperate to resolve telomeric D loops in a manner regulated by TRF1 and TRF2. *Molecular cell,* 14(6), pp. 763-774.

O'Sullivan, R.J. and Almouzni, G. (2014). Assembly of telomeric chromatin to create ALTernative endings. *Trends in cell biology,* 24(11), pp. 675-685.

Palm, W. and de Lange, T. (2008). How shelterin protects mammalian telomeres. *Annual Review of Genetics,* 42, pp. 301-334.

Park, D.J., Pask, A.J., Huynh, K., Renfree, M.B., Harley, V.R. and Graves, J.A. (2004). Comparative analysis of ATRX, a chromatin remodeling protein. *Gene,* 339, pp. 39-48.

Perlmann, T. and Jansson, L. (1995). A novel pathway for vitamin A signaling mediated by RXR heterodimerization with NGFI-B and NURR1. *Genes & development,* 9(7), pp. 769-782.

Popp, M.W. and Maquat, L.E. (2016). Leveraging Rules of Nonsense-Mediated mRNA Decay for Genome Engineering and Personalized Medicine. *Cell,* 165(6), pp. 1319-1322.

Porras, A., Gaillard, S. and Rundell, K. (1999). The simian virus 40 small-t and large-T antigens jointly regulate cell cycle reentry in human fibroblasts. *Journal of virology,* 73(4), pp. 3102-3107.

Potts, P.R. and Yu, H. (2007). The SMC5/6 complex maintains telomere length in ALT cancer cells through SUMOylation of telomere-binding proteins. *Nature structural & molecular biology,* 14(7), pp. 581-590.

Pourcel, C., Salvignol, G. and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England),* 151(Pt 3), pp. 653-663.

Qin, J., Chen, X., Yu-Lee, L.Y., Tsai, M.J. and Tsai, S.Y. (2010). Nuclear receptor COUP-TFII controls pancreatic islet tumor angiogenesis by regulating vascular endothelial growth factor/vascular endothelial growth factor receptor-2 signaling. *Cancer research,* 70(21), pp. 8812-8821.

Qin, J., Wu, S.P., Creighton, C.J., Dai, F., Xie, X., Cheng, C.M., Frolov, A., Ayala, G., Lin, X., Feng, X.H., Ittmann, M.M., Tsai, S.J., Tsai, M.J. and Tsai, S.Y. (2013). COUP-TFII inhibits TGF-beta-induced growth barrier to promote prostate tumorigenesis. *Nature,* 493(7431), pp. 236-240.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England),* 26(6), pp. 841-842.

Ramamoorthy, M. and Smith, S. (2015). Loss of ATRX Suppresses Resolution of Telomere Cohesion to Control Recombination in ALT Cancer Cells. *Cancer cell,* 28(3), pp. 357-369.

Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. and Zhang, F. (2013). Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell,* 154(6), pp. 1380-1389.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A. and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nature protocols,* 8(11), pp. 2281-2308.

Ranhotra, H.S. (2015). The NR4A orphan nuclear receptors: mediators in metabolism and diseases. *Journal of receptor and signal transduction research,* 35(2), pp. 184-188.

Rathi, A.V., Saenz Robles, M.T., Cantalupo, P.G., Whitehead, R.H. and Pipas, J.M. (2009). Simian virus 40 T-antigen-mediated gene regulation in enterocytes is controlled primarily by the Rb-E2F pathway. *Journal of virology,* 83(18), pp. 9521-9531.

Ratnakumar, K., Duarte, L.F., LeRoy, G., Hasson, D., Smeets, D., Vardabasso, C., Bonisch, C., Zeng, T., Xiang, B., Zhang, D.Y., Li, H., Wang, X., Hake, S.B., Schermelleh, L., Garcia, B.A. and Bernstein, E. (2012). ATRX-mediated chromatin association of histone variant macroH2A1 regulates alpha-globin expression. *Genes & development,* 26(5), pp. 433-438.

Reuter, S., Bartelmann, M., Vogt, M., Geisen, C., Napierski, I., Kahn, T., Delius, H., Lichter, P., Weitz, S., Korn, B. and Schwarz, E. (1998). APM-1, a novel human gene, identified by aberrant co-transcription with papillomavirus oncogenes in a cervical carcinoma cell line, encodes a BTB/POZ-zinc finger protein with growth inhibitory activity. *The EMBO journal,* 17(1), pp. 215-222.

Riggins, R.B., Mazzotta, M.M., Maniya, O.Z. and Clarke, R. (2010). Orphan nuclear receptors in breast cancer pathogenesis and therapeutic response. *Endocrine-Related Cancer,* 17, pp. R213-31.

Rippe, K. and Luke, B. (2015). TERRA and the state of the telomere. *Nature structural & molecular biology,* 22(11), pp. 853-858.

Ritchie, K., Seah, C., Moulin, J., Isaac, C., Dick, F. and Berube, N.G. (2008). Loss of ATRX leads to chromosome cohesion and congression defects. *The Journal of cell biology,* 180(2), pp. 315-324.

Robinson-Rechavi, M., Escriva Garcia, H. and Laudet, V. (2003). The nuclear receptor superfamily. *Journal of cell science,* 116(Pt 4), pp. 585-586.

Rosette, C. and Karin, M. (1995). Cytoskeletal control of gene expression: depolymerization of microtubules activates NF-kappa B. *The Journal of cell biology,* 128(6), pp. 1111-1119.

Royle, N.J., Mendez-Bermudez, A., Gravani, A., Novo, C., Foxon, J., Williams, J., Cotton, V. and Hidalgo, A. (2009). The role of recombination in telomere length maintenance. *Biochem Soc Trans,* 37, pp. 589-595.

Sablin, E.P., Woods, A., Krylova, I.N., Hwang, P., Ingraham, H.A. and Fletterick, R.J. (2008). The structure of corepressor Dax-1 bound to its target nuclear receptor LRH-1. *Proceedings of the National Academy of Sciences of the United States of America,* 105(47), pp. 18390-18395.

Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J. and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. *EMBO reports,* 16(7), pp. 836-850.

Saenz-Robles, M.T., Markovics, J.A., Chong, J.L., Opavsky, R., Whitehead, R.H., Leone, G. and Pipas, J.M. (2007). Intestinal hyperplasia induced by simian virus 40 large tumor antigen requires E2F2. *Journal of virology,* 81(23), pp. 13191-13199.

Safe, S., Jin, U.H., Hedrick, E., Reeder, A. and Lee, S.O. (2014). Minireview: role of orphan nuclear receptors in cancer and potential as drug targets. *Molecular endocrinology (Baltimore, Md.),* 28(2), pp. 157-172.

Saini, N., Ramakrishnan, S., Elango, R., Ayyar, S., Zhang, Y., Deem, A., Ira, G., Haber, J.E., Lobachev, K.S. and Malkova, A. (2013). Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature,* 502(7471), pp. 389-392.

Sakofsky, C.J. and Malkova, A. (2017). Break induced replication in eukaryotes: mechanisms, functions, and consequences. *Critical reviews in biochemistry and molecular biology,* 52(4), pp. 395-413.

Salas, T.R., Petruseva, I., Lavrik, O., Bourdoncle, A., Mergny, J.L., Favre, A. and Saintome, C. (2006). Human replication protein A unfolds telomeric G-quadruplexes. *Nucleic acids research,* 34(17), pp. 4857-4865.

Sandelin, A. and Wasserman, W.W. (2005). Prediction of nuclear hormone receptor response elements. *Molecular endocrinology (Baltimore, Md.),* 19(3), pp. 595-606.

Sander, J.D. and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology,* 32(4), pp. 347-355.

Sandin, S. and Rhodes, D. (2014). Telomerase structure. *Current opinion in structural biology,* 25, pp. 104-110.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America,* 74(12), pp. 5463-5467.

Sarma, K., Cifuentes-Rojas, C., Ergun, A., Del Rosario, A., Jeon, Y., White, F., Sadreyev, R. and Lee, J.T. (2014). ATRX directs binding of PRC2 to Xist RNA and Polycomb targets. *Cell,* 159(4), pp. 869-883.

Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tonjes, M., Hovestadt, V., Albrecht, S., Kool, M., Nantel, A., Konermann, C., Lindroth, A., Jager, N., Rausch, T., Ryzhova, M., Korbel, J.O., Hielscher, T., Hauser, P., Garami, M., Klekner, A., Bognar, L., Ebinger, M., Schuhmann, M.U., Scheurlen, W., Pekrun, A., Fruhwald, M.C., Roggendorf, W., Kramm, C., Durken, M., Atkinson, J., Lepage, P., Montpetit, A., Zakrzewska, M., Zakrzewski, K., Liberski, P.P., Dong, Z., Siegel, P., Kulozik, A.E., Zapatka, M., Guha, A., Malkin, D., Felsberg, J., Reifenberger, G., von Deimling, A., Ichimura, K., Collins, V.P., Witt, H., Milde, T., Witt, O., Zhang, C., Castelo-Branco, P., Lichter, P., Faury, D., Tabori, U., Plass, C., Majewski, J., Pfister, S.M. and Jabado, N. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature,* 482(7384), pp. 226-231.

Sentis, S., Le Romancer, M., Bianchin, C., Rostan, M.C. and Corbo, L. (2005). Sumoylation of the estrogen receptor alpha hinge region regulates its transcriptional activity. *Molecular endocrinology (Baltimore, Md.),* 19(11), pp. 2671-2684.

Sfeir, A.J., Chai, W., Shay, J.W. and Wright, W.E. (2005). Telomere-end processing the terminal nucleotides of human chromosomes. *Molecular cell,* 18(1), pp. 131-138.

Shackney, S.E., Smith, C.A., Miller, B.W., Burholt, D.R., Murtha, K., Giles, H.R., Ketterer, D.M. and Pollice, A.A. (1989). Model for the genetic evolution of human solid tumors. *Cancer research,* 49(12), pp. 3344-3354.

Shampay, J., Szostak, J.W. and Blackburn, E.H. (1984). DNA sequences of telomeres maintained in yeast. *Nature,* 310(5973), pp. 154-157.

Shay, J.W., Van Der Haegen, B.A., Ying, Y. and Wright, W.E. (1993). The frequency of immortalization of human fibroblasts and mammary epithelial cells transfected with SV40 large T-antigen. *Experimental cell research,* 209(1), pp. 45-52.

Shay, J.W. and Wright, W.E. (2005). Senescence and immortalization: role of telomeres and telomerase. *Carcinogenesis,* 26(5), pp. 867-874.

Shay, J.W. and Wright, W.E. (2000). Hayflick, his limit, and cellular ageing. *Nature Reviews Molecular Cell Biology,* 1, pp. 72-76.

Shen, R. and Seshan, V.E. (2016). FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic acids research,* 44(16), pp. e131.

Sheppard, H.M., Corneillie, S.I., Espiritu, C., Gatti, A. and Liu, X. (1999). New insights into the mechanism of inhibition of p53 by simian virus 40 large T antigen. *Molecular and cellular biology,* 19(4), pp. 2746-2753.

Shi, Y. (2009). Serine/threonine phosphatases: mechanism through structure. *Cell,* 139(3), pp. 468-484.

Shi, Y. (2007). Orphan nuclear receptors in drug discovery. *Drug discovery today,* 12(11-12), pp. 440-445.

Shibata, T., Shimoyama, Y., Gotoh, M. and Hirohashi, S. (1997). Identification of human cadherin-14, a novel neurally specific type II cadherin, by protein interaction cloning. *The Journal of biological chemistry,* 272(8), pp. 5236-5240.

Silverman, J., Takai, H., Buonomo, S.B., Eisenhaber, F. and de Lange, T. (2004). Human Rif1, ortholog of a yeast telomeric protein, is regulated by ATM and 53BP1 and functions in the S-phase checkpoint. *Genes & development,* 18(17), pp. 2108-2119.

Smogorzewska, A. and de Lange, T. (2002). Different telomere damage signaling pathways in human and mouse cells. *The EMBO Journal,* 21, pp. 4338-4348.

Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A. and de Lange, T. (2002). DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Current biology : CB,* 12(19), pp. 1635-1644.

Sobinoff, A.P., Allen, J.A., Neumann, A.A., Yang, S.F., Walsh, M.E., Henson, J.D., Reddel, R.R. and Pickett, H.A. (2017). BLM and SLX4 play opposing roles in recombination-dependent replication at human telomeres. *The EMBO journal,* 36(19), pp. 2907-2919.

Stadler, M., Chelbi-Alix, M.K., Koken, M.H., Venturini, L., Lee, C., Saib, A., Quignon, F., Pelicano, L., Guillemin, M.C. and Schindler, C. (1995). Transcriptional induction of the PML growth suppressor gene by interferons is mediated through an ISRE and a GAS element. *Oncogene,* 11(12), pp. 2565-2573.

Stagno D'Alcontres, M., Mendez-Bermudez, A., Foxon, J.L., Royle, N.J. and Salomoni, P. (2007). Lack of TRF2 in ALT cells causes PML-dependent p53 activation and loss of telomeric DNA. *The Journal of cell biology,* 179(5), pp. 855-867.

Starling, J.A., Maule, J., Hastie, N.D. and Allshire, R.C. (1990). Extensive telomere repeat arrays in mouse are hypervariable. *Nucleic acids research,* 18(23), pp. 6881-6888.

Stavropoulos, D.J., Bradshaw, P.S., Li, X., Pasic, I., Truong, K., Ikura, M., Ungrin, M. and Meyn, M.S. (2002). The Bloom syndrome helicase BLM interacts with TRF2 in ALT cells and promotes telomeric DNA synthesis. *Human molecular genetics,* 11(25), pp. 3135-3144.

Stewart, J.A., Wang, F., Chaiken, M.F., Kasbek, C., Chastain, P.D.,2nd, Wright, W.E. and Price, C.M. (2012). Human CST promotes telomere duplex replication and general replication restart after fork stalling. *The EMBO journal,* 31(17), pp. 3537-3549.

Sturzenegger, A., Burdova, K., Kanagaraj, R., Levikova, M., Pinto, C., Cejka, P. and Janscak, P. (2014). DNA2 cooperates with the WRN and BLM RecQ helicases to mediate long-range DNA end resection in human cells. *The Journal of biological chemistry,* 289(39), pp. 27314-27326.

Sun, H., Karow, J.K., Hickson, I.D. and Maizels, N. (1998). The Bloom's syndrome helicase unwinds G4 DNA. *The Journal of biological chemistry,* 273(42), pp. 27587-27592.

Sung, P., Krejci, L., Van Komen, S. and Sehorn, M.G. (2003). Rad51 recombinase and recombination mediators. *The Journal of biological chemistry,* 278(44), pp. 42729-42732.

Supek, F., Bosnjak, M., Skunca, N. and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one,* 6(7), pp. e21800.

Takai, H., Smogorzewska, A. and de Lange, T. (2003). DNA damage foci at dysfunctional telomeres. *Current biology : CB,* 13, pp. 1549-1556.

Takakura, M., Kyo, S., Kanaya, T., Hirano, H., Takeda, J., Yutsudo, M. and Inoue, M. (1999). Cloning of human telomerase catalytic subunit (hTERT) gene promoter and identification of proximal core promoter sequences essential for transcriptional activation in immortalized and cancer cells. *Cancer research,* 59(3), pp. 551-557.

Tan, M.H., Zhou, X.E., Soon, F.F., Li, X., Li, J., Yong, E.L., Melcher, K. and Xu, H.E. (2013). The crystal structure of the orphan nuclear receptor NR2E3/PNR ligand binding domain reveals a dimeric auto-repressed conformation. *PloS one,* 8(9), pp. e74359.

Tang, J., Wu, S., Liu, H., Stratt, R., Barak, O.G., Shiekhattar, R., Picketts, D.J. and Yang, X. (2004). A novel transcription regulatory complex containing death domain-associated protein and the ATR-X syndrome protein. *The Journal of biological chemistry,* 279(19), pp. 20369-20377.

Tarsounas, M., Munoz, P., Claas, A., Smiraldo, P.G., Pittman, D.L., Blasco, M.A. and West, S.C. (2004). Telomere maintenance requires the RAD51D recombination/repair protein. *Cell,* 117(3), pp. 337-347.

Tarsounas, M. and Tijsterman, M. (2013). Genomes and G-quadruplexes: for better or for worse. *Journal of Molecular Biology,* 425(23), pp. 4782-4789.

Teasley, D.C., Parajuli, S., Nguyen, M., Moore, H.R., Alspach, E., Lock, Y.J., Honaker, Y., Saharia, A., Piwnica-Worms, H. and Stewart, S.A. (2015). Flap Endonuclease 1 Limits Telomere Fragility on the Leading Strand. *The Journal of biological chemistry,* 290(24), pp. 15133-15145.

Teng, S.C. and Zakian, V.A. (1999). Telomere-telomere recombination is an efficient bypass pathway for telomere maintenance in Saccharomyces cerevisiae. *Molecular and cellular biology,* 19(12), pp. 8083-8093.

Thompson, J. and Winoto, A. (2008). During negative selection, Nur77 family proteins translocate to mitochondria where they associate with Bcl-2 and expose its proapoptotic BH3 domain. *The Journal of experimental medicine,* 205(5), pp. 1029-1036.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L. and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols,* 7(3), pp. 562-578.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J. and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology,* 28(5), pp. 511-515.

Udugama, M., Sanij, E., Voon, H.P.J., Son, J., Hii, L., Henson, J.D., Chan, F.L., Chang, F.T.M., Liu, Y., Pearson, R.B., Kalitsis, P., Mann, J.R., Collas, P., Hannan, R.D. and Wong, L.H. (2018). Ribosomal DNA copy loss and repeat instability in ATRX-mutated cancers. *Proceedings of the National Academy of Sciences of the United States of America,* 115(18), pp. 4737-4742.

Uringa, E.J., Lisaingo, K., Pickett, H.A., Brind'Amour, J., Rohde, J.H., Zelensky, A., Essers, J. and Lansdorp, P.M. (2012). RTEL1 contributes to DNA replication and repair and telomere maintenance. *Molecular biology of the cell,* 23(14), pp. 2782-2792.

van Brabant, A.J., Ye, T., Sanz, M., German III, J.L., Ellis, N.A. and Holloman, W.K. (2000). Binding and melting of D-loops by the Bloom syndrome helicase. *Biochemistry,* 39(47), pp. 14617-14625.

van Hamburg, J.P., de Bruijn, M.J., Dingjan, G.M., Beverloo, H.B., Diepstraten, H., Ling, K.W. and Hendriks, R.W. (2008). Cooperation of Gata3, c-Myc and Notch in malignant transformation of double positive thymocytes. *Molecular immunology,* 45(11), pp. 3085-3095.

Vannier, J.B., Pavicic-Kaltenbrunner, V., Petalcorin, M.I., Ding, H. and Boulton, S.J. (2012). RTEL1 dismantles T loops and counteracts telomeric G4-DNA to maintain telomere integrity. *Cell,* 149(4), pp. 795-806.

Varley, H., Pickett, H.A., Foxon, J.L., Reddel, R.R. and Royle, N.J. (2002). Molecular characterization of inter-telomere and intra-telomere mutations in human ALT cells. *Nature Genetics,* 30, pp. 301-305.

Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M. and Ng, P.C. (2016). SIFT missense predictions for genomes. *Nature protocols,* 11(1), pp. 1-9.

Vessey, A.R. and Culp, L.A. (1978). Contact-inhibited revertant cell lines isolated from SV40-transformed cells. VIII. Membrane protein and glycoprotein composition. *Virology,* 86(2), pp. 556-561.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y. and Fuks, F. (2006). The Polycomb group protein EZH2 directly controls DNA methylation. *Nature,* 439(7078), pp. 871-874.

Vouillot, L., Thelie, A. and Pollet, N. (2015). Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. *G3 (Bethesda, Md.),* 5(3), pp. 407-415.

Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S. and Vidal, M. (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods in enzymology,* 328, pp. 575-592.

Wang, F., Stewart, J. and Price, C.M. (2014). Human CST abundance determines recovery from diverse forms of DNA damage and replication stress. *Cell cycle (Georgetown, Tex.),* 13(22), pp. 3488-3498.

Wang, K., Li, M. and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research,* 38(16), pp. e164.

Wang, R.C., Smogorzewska, A. and de Lange, T. (2004). Homologous recombination generates T-loop-sized deletions at human telomeres. 119, pp. 355-368.

Wang, Z., Benoit, G., Liu, J., Prasad, S., Aarnisalo, P., Liu, X., Xu, H., Walker, N.P. and Perlmann, T. (2003). Structure and function of Nurr1 identifies a class of ligand-independent nuclear receptors. *Nature,* 423(6939), pp. 555-560.

Wilson, T.E., Fahrner, T.J., Johnston, M. and Milbrandt, J. (1991). Identification of the DNA binding site for NGFI-B by genetic selection in yeast. *Science (New York, N.Y.),* 252(5010), pp. 1296-1300.

Wilson, T.E., Fahrner, T.J. and Milbrandt, J. (1993). The orphan receptors NGFI-B and steroidogenic factor 1 establish monomer binding as a third paradigm of nuclear receptor-DNA interaction. *Molecular and cellular biology,* 13(9), pp. 5794-5804.

Wong, L.H., McGhie, J.D., Sim, M., Anderson, M.A., Ahn, S., Hannan, R.D., George, A.J., Morgan, K.A., Mann, J.R. and Choo, K.H. (2010). ATRX interacts with H3.3 in maintaining telomere structural integrity in pluripotent embryonic stem cells. *Genome research,* 20(3), pp. 351-360.

Wong, M. and Fish, E.N. (1998). RANTES and MIP-1alpha activate stats in T cells. *The Journal of biological chemistry,* 273(1), pp. 309-314.

Wood, R.D. and Doublie, S. (2016). DNA polymerase theta (POLQ), double-strand break repair, and cancer. *DNA repair,* 44, pp. 22-32.

Wright, W.E., Pereira-Smith, O.M. and Shay, J.W. (1989). Reversible cellular senescence: implications for immortalization of normal human diploid fibroblasts. *Molecular and cellular biology,* 9(7), pp. 3088-3092.

Wright, W.E. and Shay, J.W. (1992). The two-stage mechanism controlling cellular senescence and immortalization. *Experimental gerontology,* 27(4), pp. 383-389.

Wu, G., Lee, W.H. and Chen, P.L. (2000). NBS1 and TRF1 colocalize at promyelocytic leukemia bodies during late S/G2 phases in immortalized telomerase-negative cells. Implication of NBS1 in alternative lengthening of telomeres. *The Journal of biological chemistry,* 275(39), pp. 30618-30622.

Wu, Q. and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell,* 97(6), pp. 779-790.

Wu, Y., Shin-ya, K. and Brosh, R.M.,Jr (2008). FANCJ helicase defective in Fanconia anemia and breast cancer unwinds G-quadruplex DNA to defend genomic stability. *Molecular and cellular biology,* 28(12), pp. 4116-4128.

Xin, H., Liu, D., Wan, M., Safari, A., Kim, H., Sun, W., O'Connor, M.S. and Songyang, Z. (2007). TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase. *Nature,* 445(7127), pp. 559-562.

Xing, J., Sheppard, H.M., Corneillie, S.I. and Liu, X. (2001). p53 Stimulates TFIID-TFIIA-promoter complex assembly, and p53-T antigen complex inhibits TATA binding protein-TATA interaction. *Molecular and cellular biology,* 21(11), pp. 3652-3661.

Yan, Z.H., Karam, W.G., Staudinger, J.L., Medvedev, A., Ghanayem, B.I. and Jetten, A.M. (1998). Regulation of peroxisome proliferator-activated receptor alpha-induced transactivation by the nuclear orphan receptor TAK1/TR4. *The Journal of biological chemistry,* 273(18), pp. 10948-10957.

Yano, K., Morotomi-Yano, K., Wang, S.Y., Uematsu, N., Lee, K.J., Asaithamby, A., Weterings, E. and Chen, D.J. (2008). Ku recruits XLF to DNA double-strand breaks. *EMBO reports,* 9(1), pp. 91-96.

Yeager, T.R., Neumann, A.A., Englezou, A., Huschtscha, L.I., Noble, J.R. and Reddel, R.R. (1999). Telomerase-negative immortalized human cells contain a novel type of promyelocytic leukemia (PML) body. *Cancer Research,* 59, pp. 4175-4179.

Yin, K., Chhabra, Y., Tropee, R., Lim, Y.C., Fane, M., Dray, E., Sturm, R.A. and Smith, A.G. (2017). NR4A2 Promotes DNA Double-strand Break Repair Upon Exposure to UVR. *Molecular cancer research : MCR,* 15(9), pp. 1184-1196.

Yoon, K., Lee, S.O., Cho, S.D., Kim, K., Khan, S. and Safe, S. (2011). Activation of nuclear TR3 (NR4A1) by a diindolylmethane analog induces apoptosis and proapoptotic genes in pancreatic cancer cells and tumors. *Carcinogenesis,* 32(6), pp. 836-842.

Zaug, A.J., Podell, E.R. and Cech, T.R. (2005). Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proceedings of the National Academy of Sciences of the United States of America,* 102(31), pp. 10864-10869.

Zhang, D.H., Zhou, B., Huang, Y., Xu, L.X. and Zhou, J.Q. (2006). The human Pif1 helicase, a potential Escherichia coli RecD homologue, inhibits telomerase activity. *Nucleic acids research,* 34(5), pp. 1393-1404.

Zhang, Y., Mao, D., Roswit, W.T., Jin, X., Patel, A.C., Patel, D.A., Agapov, E., Wang, Z., Tidwell, R.M., Atkinson, J.J., Huang, G., McCarthy, R., Yu, J., Yun, N.E., Paessler, S., Lawson, T.G., Omattage, N.S., Brett, T.J. and Holtzman, M.J. (2015). PARP9-DTX3L ubiquitin ligase targets host histone H2BJ and viral 3C protease to

enhance interferon signaling and control viral infection. *Nature immunology,* 16(12), pp. 1215-1227.

Zhao, D., Desai, S. and Zeng, H. (2011). VEGF stimulates PKD-mediated CREB-dependent orphan nuclear receptor Nurr1 expression: role in VEGF-induced angiogenesis. *International journal of cancer,* 128(11), pp. 2602-2612.

Zhong, Z.H., Jiang, W.Q., Cesare, A.J., Neumann, A.A., Wadhwa, R. and Reddel, R.R. (2007). Disruption of telomere maintenance by depletion of the MRE11/RAD50/NBS1 complex in cells that use alternative lengthening of telomeres. *The Journal of biological chemistry,* 282(40), pp. 29314-29322.

Zimmermann, M. and de Lange, T. (2014). 53BP1: pro choice in DNA repair. *Trends in cell biology,* 24(2), pp. 108-117.

Appendixes

Appendixes - chapter 1

Appendix 1.1 ALT related genes

Gene Symbol	Validation	Description	Reference
BLM	LKE	RecQ helicases	Bhattacharyya et al. 2009; Sobinoff et al 2017
FANCA	K	Fanconi Anaemia	Fan et al., 2009
FANCD2	K	Fanconi Anaemia	Fan et al., 2009
FEN1	K	Endonuclease	Saharia and Stewart, 2009
MRE11A	K	Subunit of MRN complex	Zhong et al., 2007, Henson et al., 2009
MUS81	K	Endonuclease	Zeng et al. 2009
NBN	K	NBS1 protein; Subunit of MRN complex	Zhong et al., 2007
NSMCE2 / MMS21	K	MMS21 Homolog; SMC5-SMC6 Complex	Potts & Yu, 2007
PCNA	K	Replication; RFC–PCNA functions as a telomere damage sensor	Jiang et al. 2009; Garcia-Exposito et al. 2016; Dilley et al. 2016
RAD50	LK	Subunit of MRN complex	Zhong et al., 2007, Henson et al., 2009
RAD51A	LK	RecA-like protein; a ss DNA-binding protein that promotes homology-directed searches and strand invasion at ALT telomeres	Draskovic et al. 2009; Cho et al., 2014
RAD51C	LK	RecA-like protein; a ss DNA-binding protein that promotes homology-directed searches and strand invasion at ALT telomeres	Draskovic et al. 2009; Cho et al., 2014
RAD51D	LK	RecA-like protein; a ss DNA-binding protein that promotes homology-directed searches and strand invasion at ALT telomeres	Draskovic et al. 2009; Cho et al., 2014
SMC5	K	SMC5-SMC6 Complex	Potts & Yu, 2007
SMC6	K	SMC5-SMC6 Complex	Potts & Yu, 2007
TERF2	LK E	TRF2 protein; Shelterin Complex Subunit ; ds telomere DNA binding; fusion; T-circles	Yeager et al 1999; Jiang et al. 2007; Stagno D'Alcontres et al. 2007
TOP3A	K	DNA Topoisomerase III Alpha	Raynard et al. 2006; Temime-Smaali et al. 2008
SMARCAL1	LK	ATP-dependent DNA- helicase	Poole et al. 2015; Cox et al. 2016
POT1	L	Shelterin Complex Subunit	Denchi et al. 2007
TERF1	LK	TRF1 protein; Shelterin Complex Subunit	Yeager et al 1999; Jiang et al. 2007
TINF2	LK	TIN2 protein; Shelterin Complex Subunit,	Jiang et al. 2007; Temime-Smaali et al. 2008

CTC1	K	Subunit of CTC1-STN1-TEN1 complex	Huang et al 2017
STN1	K	Subunit of CTC1-STN1-TEN1 complex	Huang et al 2017
TEN1	K	Subunit of CTC1-STN1-TEN1 complex	Huang et al 2017
TOP2A	L	DNA Topoisomerase II Alpha	Bhattacharyya et al., 2009
FANCM	LK	Fanconi Anemia	Xiaolei et al 2017
ACD	L	TPP1 protein, Shelterin Complex Subunit	Hu et al. 2016
ATM	L	Serine/Threonine Kinase	Stagno D'Alcontres et al. 2007
ATR	L	Serine/Threonine Kinase	Barr et al. 2003
BRCA2	L	DNA damage sensor	Wu et al. 2003; Acharya et al 2014
BRIP1/ FANCJ	L	structure-specific DNA helicase	Déjardin et al. 2009
CBX1/ HP1β	L	Heterochromatin Protein 1 Homolog Beta	Jiang et al. 2007
CBX3/ HP1γ	L	Heterochromatin Protein 1 Homolog gamma	Jiang et al. 2007
CBX5/ HP1α	L	Heterochromatin Protein 1 Homolog Alpha	Jiang et al. 2007
CDK2	L	Cell cycle regulator	Wu 2003
ERCC1	L	endonuclease; Excision Repair	Zhu et al 2003
ERCC4/XPF	L	endonuclease; Excision Repair, T-circles	Sobinoff et al 2017
FANCL	K	Fanconi Anaemia	Fan et al., 2009
H2AFX	L	DNA damage response protien; TIFs	Nabetani et al 2004; Cesare et al 2009
HUS1	L	Component of the 9-1-1 (RAD9-RAD1-HUS1) complex; checkpoint complex	Nabetani et al. 2004
MDC1	L	Mediator of DNA Damage Checkpoint 1	Cesare et al 2009
MORC3/NXP2	L	MORC Family CW-Type Zinc Finger 3	Osterwald et al 2015
PARP2	L	poly(ADP-Ribose) polymerase 2 (PARP-2); DSBR	Dantzer et al. 2004
PML	L	Member of the tripartite motif (TRIM) family; TF; PML-nuclear bodies	Yeager et al. 1999; Jiang et al. 2007
RAD1	L	Subunit of RAD9-RAD1-HUS1 complex	Nabetani et al. 2004
RAD17	L	Subunit of RAD9-RAD1-HUS1 complex	Nabetani et al. 2004
RAD52	L	Subunit of RAD9-RAD1-HUS1 complex	Lundblad & Blackburn 1993; Yeager et al 1999; Min et al 2017
RAD9A	L	Subunit of RAD9-RAD1-HUS1 complex	Nabetani et al. 2004
RAD9B	L	Subunit of RAD9-RAD1-HUS1 complex	Nabetani et al. 2004

RAP1A	L	Shelterin Complex Subunit ; TRF2-interacting	Jiang et al. 2007
RIF1	L	RAP1 Interacting Factor Homolog; DDR for DSB	Silverman et al. 2004; Escribano-Díaz et al 2013
RPA1	L	Subunit of the heterotrimeric RPA complex	Grudic et al. 2007
RPA2	L	Subunit of the heterotrimeric RPA complex	Grudic et al. 2007
TEP1	L	Telomerase-associated protein 1	Bhattacharyya et al., 2009
WRN	LK	RecQ-like helicase and nuclease	Johnson et al. 2001; Mendez-Bermudez et al. 2012; Gocha et al. 2014
XRCC3	K	Member of the RecA/Rad51-related protein family; T-circles	Compton et al. 2017
RTEL1	LK	ATP-dependent DNA helicase implicated in telomere-length regulation	Vannier et al 2012
PAXIP1	P	DNA damage response protein; **New ALT candidate**	Chu et al. 2017
TP53BP1	LK	53BP1 protein; DNA damage response; TIFs	Nabetani et al 2004; Jiang et al. 2007; Cesare et al 2009
PSMC3IP/HOP2	LK	Subunit of the PSMC3IP/MND1 complex	Cho et al., 2014
MND1	LK	Subunit of the PSMC3IP/MND1 complex	Cho et al., 2014
HSP90AA1	L	Heat Shock Protein HSP 90-Alpha	Bhattacharyya et al. 2009
HSP90AB1	L	Heat Shock Protein HSP 90-Beta	Bhattacharyya et al. 2009
ASF1A	LK	Anti-Silencing Function 1A Histone Chaperone	O'Sullivan et al. 2014
ASF1B	LK	Anti-Silencing Function 1B Histone Chaperone	O'Sullivan et al. 2014
ATRX	LK	Chromatin Remodeler	Lovejoy et al. 2012
BRCA1	L	DNA damage sensor protein	Wu et al. 2003; Kargaran et al 2016
DAXX	LK	Multifunctional protein; Transcriptional factor; ATRX partner	Lovejoy et al. 2012
H3F3A	S	Histone variant H3.3a	Lovejoy et al. 2012
H3F3B	-	Histone variant H3.3b	Lovejoy et al. 2012
SP100	LKE	Major component of the PML bodies	Jiang et al. 2005
TP53	LE	Tumour suppression; Transcription factor	Chen et al. 2006; Stagno D'Alcontres et al. 2007
SLX4	LKE	Structure-Specific Endonuclease Subunit; T-circles	Vannier et al 2012; Sobinoff et al 2017

ALT related genes. Validation Key= L = localisation, K= knockdown, E= ectopic expression.

Appendixes - chapter 2

APPENDIX 2.1 Oligonucleotide sequences

Name	5' > 3' sequence
Chromosome/ Allele specific STELA	
Telorette 2	TGCCCGTGCATCTGGCATCTAACCCT
Teltail	TGCTCCGTGCATCTGGCATC
12qSTELA	CGAAGCAGCATTCTCCTCAG
XpYpE2	TTGTCTCAGGGTCCTAGTG
17p6	GGCTGAACTATAGCCTCTGC
XpYp-427G/415C	GGTTATCGACCAGGTGCTCC
XpYp-427A/415T	GGTTATCAACCAGGTGCTCT
Fusion PCR and sequencing	
16p1	TGGACTTCTCACTTCTAGGGCAG
16p1nest	CAAGTGAGGCTCTGGGTTCC
16pTel1.1R	ATTATAGGGAAACACCCGGA
16pTel1.5F	AGGAACAGCAAAGGAAATCC
16pTel1.6F	GGAAGTCTGGGCTAAGAGAC
16pTel2.3F	GCTCTGGAGACCTGATGCTG
16pTel3F	CCACGAGGGCTCAGATTCAG
16pTel4F	GAGGTGCCATCGACTTGGAG
16pTel5.1F	GATGATGCCTGCTGTGAACG
Tsk8M	GACCAGGTTTTCCAGTGTGT
Tsk8Mnest	TTCTTCACCAGCACTCTGAG
TSk8Y	TCATTCCCGACAGGGCTCAGC
XpYp E2	TTGTCTCAGGGTCCTAGTG
CRISPR Vectors cloning and sequencing	
ATRX_PG1-F	GTCTACGCAACCTTGGTCGAAGTTTTAGAGCTAGAAATAGCA
ATRX_PG1-R	TTCGACCAAGGTTGCGTAGACGGTGTTTCGTCCTTTCCACA
ATRX-PG2-F	GGGGTGTAGTCTTTACACGTGGTTTTAGAGCTAGAAATAGCA
ATRX-PG2-F	CACGTGTAAAGACTACACCCCGGTGTTTCGTCCTTTCCACA
PX459-CS-F	GTTTCGCCACCTCTGACTTG
PX459-CS-R	GTCATGTACTGGGCACAATG
PX459-PuroF	TGGAGAGGGCAGAGGAAGTC
PX459-PuroR	GGCAACTAGAAGGCACAGTCG
PX459-Cas9F	TTACCTGGAGCACCTGCCTG
ATRX-Exon9 PCR primers	
ATRX EX9a.F	TGAGTTTTGTTGGGAATAATGTTT
ATRX EX9a.R	ATCGCTCGAAACTCGGAAT
ATRX EX9c.F	ATGGCAGCAGTGGAACTGA

ATRX EX9c2.R	CCACAGTCTCTGATTGCTTAG

APPENDIX 2.2 List of primary and secondary antibodies

Antibody	Type	Species specificity	Host species	Supplier	Cat #	Application/dilution	
						WB	IF
NR2C2	Primary	Human	Mouse	Perseus Proteomics	PP-H0107B-00	1/500	1/100
NR2F2	Primary	Human	Mouse	Perseus Proteomics	PP-H7147-00	1/500	1/100
H2A.X (Ser139)	Primary	Human	Mouse	Millipore	05-636	-	1/1000
GAPDH	Primary	Human	Mouse	Ambion	AM4300	1/10,000	-
ATRX-H-300	Primary	Human	Rabbit	Santa Cruz	sc-15408	1/300	1/100
PML	Primary	Human	Rabbit	Santa Cruz	sc-966	-	1/100
β-tubulin	Primary	Human	Rabbit	abcam	ab6046	1/10,000	-
Anti IgG/HRP	Secondary	Rabbit	Donkey	GE Healthcare	NA934	1/10,000	-
Anti IgG/HRP	Secondary	Mouse	Sheep	GE Healthcare	NA931	1/10,000	-
Anti IgG/ AlexaFluor 488	Secondary	Mouse	Goat	abcam	ab150113	-	1/1000
Anti IgG/ AlexaFluor 488	Secondary	Rabbit	Donkey	abcam	ab150073	-	1/1000

List of primary and secondary antibodies that used in Western blotting (WB), immunofluorescence (IF). IgG=immunoglobulin G, HRP=horseradish peroxidase.

APPENDIX 2.3 Maps of the CRISPR-Cas9 Vectors

A)

B)

C)
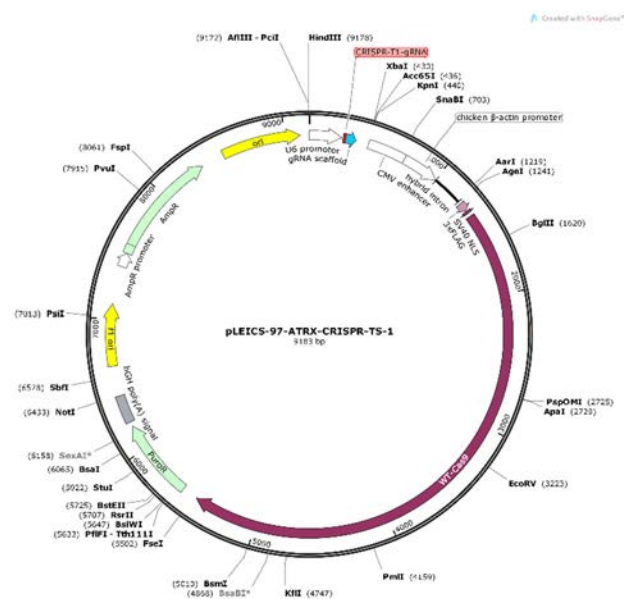
D)

Schematic of CRISPR-Cas9 vectors. (A) Addgene-PX459_v2.0, was used as the control 'empty' vector. (B) RFP1-pcDNA3.1/Zeo mammalian expression vector, was used to estimate the delivery efficiency in the CCD18LU/TCL1 cell line. (C) The pLEIC-97-ATRX-CRISPR-TS-1 produced after cloning the guide RNA for targeting ATRX exon9 into the pLEIC-97 vector. (D) The pLEIC-97-ATRX-CRISPR-TS-2 produce after cloning the guide RNA for targeting ATRX exon9 into the pLEIC-97 vector. The maps highlight all unique restriction sites.

Appendixes - chapter 4

APPENDIX 4.1 NR2F2 target genes

Gene symobl	NCBI ID	ENSEMBL ID	Full gene name
AATF	26574	ENSG00000275700	apoptosis antagonizing transcription factor
ACOT4	122970	ENSG00000177465	acyl-CoA thioesterase 4
ADAMTSL5	339366	ENSG00000185761	ADAMTS like 5
ADH5	128	ENSG00000197894	alcohol dehydrogenase 5 (class III), chi polypeptide
ADPRHL2	54936	ENSG00000116863	ADP-ribosylhydrolase like 2
ADRA1B	147	ENSG00000170214	adrenoceptor alpha 1B
ANGPTL2	23452	ENSG00000136859	angiopoietin like 2
ANKRD11	29123	ENSG00000167522	ankyrin repeat domain 11
ANKRD24	170961	ENSG00000089847	ankyrin repeat domain 24
AP2A1	160	ENSG00000196961	adaptor related protein complex 2 alpha 1 subunit
AP3B2	8120	ENSG00000103723	adaptor related protein complex 3 beta 2 subunit
ARHGAP23	57636	ENSG00000275832	Rho GTPase activating protein 23
ARHGAP4	393	ENSG00000089820	Rho GTPase activating protein 4
ARHGDIB	397	ENSG00000111348	Rho GDP dissociation inhibitor beta
ARHGEF3	50650	ENSG00000163947	Rho guanine nucleotide exchange factor 3
ARR3	407	ENSG00000120500	arrestin 3
ASGR1	432	ENSG00000141505	asialoglycoprotein receptor 1
ASXL1	171023	ENSG00000171456	additional sex combs like 1, transcriptional regulator
ATAD3C	219293	ENSG00000215915	ATPase family, AAA domain containing 3C
ATP10B	23120	ENSG00000118322	ATPase phospholipid transporting 10B (putative)
ATP1B4	23439	ENSG00000101892	ATPase Na+/K+ transporting family member beta 4
ATP8A2	51761	ENSG00000132932	ATPase phospholipid transporting 8A2
ATXN1L	342371	ENSG00000224470	ataxin 1 like
B4GALT7	11285	ENSG00000027847	beta-1,4-galactosyltransferase 7
BCL7C	9274	ENSG00000099385	BCL tumor suppressor 7C
BFSP2	8419	ENSG00000170819	beaded filament structural protein 2
BOLA1	51027	ENSG00000178096	bolA family member 1
BRD7P3	23629	ENSG00000169075	NA

C10orf62	414157	ENSG00000203942	chromosome 10 open reading frame 62
C10orf76	79591	ENSG00000120029	chromosome 10 open reading frame 76
C15orf40	123207	ENSG00000169609	chromosome 15 open reading frame 40
C16orf92	146378	ENSG00000167194	chromosome 16 open reading frame 92
C1orf61	10485	ENSG00000125462	chromosome 1 open reading frame 61
C20orf196	149840	ENSG00000171984	chromosome 20 open reading frame 196
C20orf85	128602	ENSG00000124237	chromosome 20 open reading frame 85
C2orf72	257407	ENSG00000204128	chromosome 2 open reading frame 72
C6orf10	10665	ENSG00000204296	chromosome 6 open reading frame 10
C6orf62	81688	ENSG00000112308	chromosome 6 open reading frame 62
CBR1	873	ENSG00000159228	carbonyl reductase 1
CC2D2B	387707	ENSG00000188649	coiled-coil and C2 domain containing 2B
CCDC58	131076	ENSG00000160124	coiled-coil domain containing 58
CCDC6	8030	ENSG00000108091	coiled-coil domain containing 6
CCL19	6363	ENSG00000172724	C-C motif chemokine ligand 19
CCNJL	79616	ENSG00000135083	cyclin J like
CCT3	7203	ENSG00000163468	chaperonin containing TCP1 subunit 3
CD209	30835	ENSG00000090659	CD209 molecule
CDKL3	51265	ENSG00000006837	cyclin dependent kinase like 3
CEACAM18	729767	ENSG00000213822	carcinoembryonic antigen related cell adhesion molecule 18
CEMP1	752014	ENSG00000205923	cementum protein 1
CHCHD10	400916	ENSG00000250479	coiled-coil-helix-coiled-coil-helix domain containing 10
CHPT1	56994	ENSG00000111666	choline phosphotransferase 1
CLDN10	9071	ENSG00000134873	claudin 10
CLEC4G	339390	ENSG00000182566	C-type lectin domain family 4 member G
CLEC4M	10332	ENSG00000104938	C-type lectin domain family 4 member M
CLSTN3	9746	ENSG00000139182	calsyntenin 3
CNDP1	84735	ENSG00000150656	carnosine dipeptidase 1
CNOT8	9337	ENSG00000155508	CCR4-NOT transcription complex subunit 8
COL10A1	1300	ENSG00000123500	collagen type X alpha 1 chain
COL8A2	1296	ENSG00000171812	collagen type VIII alpha 2 chain
CSF2RB	1439	ENSG00000100368	colony stimulating factor 2 receptor beta common subunit
CTSH	1512	ENSG00000103811	cathepsin H
CUL5	8065	ENSG00000166266	cullin 5
CXCL3	2921	ENSG00000163734	C-X-C motif chemokine ligand 3
CYP2E1	1571	ENSG00000130649	cytochrome P450 family 2 subfamily E member 1
CYP4F11	57834	ENSG00000171903	cytochrome P450 family 4 subfamily F member 11
CYP4Z1	199974	ENSG00000186160	cytochrome P450 family 4 subfamily Z member 1
CYP8B1	1582	ENSG00000180432	cytochrome P450 family 8 subfamily B member 1
DHDDS	79947	ENSG00000117682	dehydrodolichyl diphosphate synthase subunit
DHX16	8449	ENSG00000204560	DEAH-box helicase 16
DLGAP1	9229	ENSG00000170579	DLG associated protein 1
DLGAP3	58512	ENSG00000116544	DLG associated protein 3
DMGDH	29958	ENSG00000132837	dimethylglycine dehydrogenase

DMRTC1	63947	ENSG00000269502	DMRT like family C1
DMRTC1B	728656	ENSG00000184911	DMRT like family C1B
DNAH10	196385	ENSG00000197653	dynein axonemal heavy chain 10
DNAH2	146754	ENSG00000183914	dynein axonemal heavy chain 2
DOC2A	8448	ENSG00000149927	double C2 domain alpha
EGFL7	51162	ENSG00000172889	EGF like domain multiple 7
ELOVL5	60481	ENSG00000012660	ELOVL fatty acid elongase 5
ENPP7	339221	ENSG00000182156	ectonucleotide pyrophosphatase/phosphodiesterase 7
EPN3	55040	ENSG00000049283	epsin 3
FAM162A	26355	ENSG00000114023	family with sequence similarity 162 member A
FAM181A	90050	ENSG00000140067	family with sequence similarity 181 member A
FAM46B	115572	ENSG00000158246	family with sequence similarity 46 member B
FANCE	2178	ENSG00000112039	Fanconi anemia complementation group E
FBXL19	54620	ENSG00000099364	F-box and leucine rich repeat protein 19
FBXO2	26232	ENSG00000116661	F-box protein 2
FBXO44	93611	ENSG00000132879	F-box protein 44
FBXO6	26270	ENSG00000116663	F-box protein 6
FCER2	2208	ENSG00000104921	Fc fragment of IgE receptor II
FHAD1	114827	ENSG00000142621	forkhead associated phosphopeptide binding domain 1
FKBP8	23770	ENSG00000105701	FK506 binding protein 8
FLJ34503	285759	ENSG00000175967	uncharacterized FLJ34503
FLJ45513	729220	ENSG00000204584	uncharacterized LOC729220
FSTL1	11167	ENSG00000163430	follistatin like 1
GALK1	2584	ENSG00000108479	galactokinase 1
GALNT10	55568	ENSG00000164574	polypeptide N-acetylgalactosaminyltransferase 10
GBGT1	26301	ENSG00000148288	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1 (FORS blood group)
GDF7	151449	ENSG00000143869	growth differentiation factor 7
GDF9	2661	ENSG00000164404	growth differentiation factor 9
GDPD5	81544	ENSG00000158555	glycerophosphodiester phosphodiesterase domain containing 5
GLOD4	51031	ENSG00000167699	glyoxalase domain containing 4
GLS2	27165	ENSG00000135423	glutaminase 2
GLYAT	10249	ENSG00000149124	glycine-N-acyltransferase
GPNMB	10457	ENSG00000136235	glycoprotein nmb
GPR12	2835	ENSG00000132975	G protein-coupled receptor 12
GPR146	115330	ENSG00000164849	G protein-coupled receptor 146
GPR32	2854	ENSG00000142511	G protein-coupled receptor 32
GPR45	11250	ENSG00000135973	G protein-coupled receptor 45
GPSM2	29899	ENSG00000121957	G-protein signaling modulator 2
GPT2	84706	ENSG00000166123	glutamic--pyruvic transaminase 2
GRAMD1C	54762	ENSG00000178075	GRAM domain containing 1C
GRHL3	57822	ENSG00000158055	grainyhead like transcription factor 3
GSTM5	2949	ENSG00000134201	glutathione S-transferase mu 5
GULP1	51454	ENSG00000144366	GULP, engulfment adaptor PTB domain containing 1
GYS1	2997	ENSG00000104812	glycogen synthase 1

HAO2	51179	ENSG00000116882	hydroxyacid oxidase 2
HARS	3035	ENSG00000170445	histidyl-tRNA synthetase
HAUS6	54801	ENSG00000147874	HAUS augmin like complex subunit 6
HCG9	10255	ENSG00000204625	HLA complex group 9 (non-protein coding)
HCRT	3060	ENSG00000161610	hypocretin neuropeptide precursor
HES3	390992	ENSG00000173673	hes family bHLH transcription factor 3
HEXA	3073	ENSG00000213614	hexosaminidase subunit alpha
HIRIP3	8479	ENSG00000149929	HIRA interacting protein 3
HIST1H1D	3007	ENSG00000124575	histone cluster 1 H1 family member d
HIST1H2BH	8345	ENSG00000275713	histone cluster 1 H2B family member h
HIST1H3F	8968	ENSG00000277775	histone cluster 1 H3 family member f
HIST1H4F	8361	ENSG00000274618	histone cluster 1 H4 family member f
HIST1H4G	8369	ENSG00000275663	histone cluster 1 H4 family member g
HIST2H2AB	317772	ENSG00000184270	histone cluster 2 H2A family member b
HIST2H2BE	8349	ENSG00000184678	histone cluster 2 H2B family member e
HLA-C	3107	ENSG00000204525	major histocompatibility complex, class I, C
HMGCS2	3158	ENSG00000134240	3-hydroxy-3-methylglutaryl-CoA synthase 2
HMOX1	3162	ENSG00000100292	heme oxygenase 1
HNRNPA2B1	3181	ENSG00000122566	heterogeneous nuclear ribonucleoprotein A2/B1
HPS6	79803	ENSG00000166189	HPS6, biogenesis of lysosomal organelles complex 2 subunit 3
HSPG2	3339	ENSG00000142798	heparan sulfate proteoglycan 2
HYI	81888	ENSG00000178922	hydroxypyruvate isomerase (putative)
ICMT	23463	ENSG00000116237	isoprenylcysteine carboxyl methyltransferase
INCA1	388324	ENSG00000196388	inhibitor of CDK, cyclin A1 interacting protein 1
ING2	3622	ENSG00000168556	inhibitor of growth family member 2
ING4	51147	ENSG00000111653	inhibitor of growth family member 4
INO80D	54891	ENSG00000114933	INO80 complex subunit D
INTS7	25896	ENSG00000143493	integrator complex subunit 7
IQCD	115811	ENSG00000166578	IQ motif containing D
IRF2	3660	ENSG00000168310	interferon regulatory factor 2
IRS4	8471	ENSG00000133124	insulin receptor substrate 4
ISM1	140862	ENSG00000101230	isthmin 1
ITIH3	3699	ENSG00000162267	inter-alpha-trypsin inhibitor heavy chain 3
JAM3	83700	ENSG00000166086	junctional adhesion molecule 3
JDP2	122953	ENSG00000140044	Jun dimerization protein 2
KAZALD1	81621	ENSG00000107821	Kazal type serine peptidase inhibitor domain 1
KCNH4	23415	ENSG00000089558	potassium voltage-gated channel subfamily H member 4
KCNJ10	3766	ENSG00000177807	potassium voltage-gated channel subfamily J member 10
KCNJ16	3773	ENSG00000153822	potassium voltage-gated channel subfamily J member 16
KCNJ9	3765	ENSG00000162728	potassium voltage-gated channel subfamily J member 9
KCNK12	56660	ENSG00000184261	potassium two pore domain channel subfamily K member 12
KCTD19	146212	ENSG00000168676	potassium channel tetramerization domain containing 19
KCTD5	54442	ENSG00000167977	potassium channel tetramerization domain containing 5
KIAA0040	9674	ENSG00000235750	KIAA0040

KIR3DX1	90011	ENSG00000104970	killer cell immunoglobulin like receptor, three Ig domains X1
KLB	152831	ENSG00000134962	klotho beta
KLF9	687	ENSG00000119138	Kruppel like factor 9
KLHL25	64410	ENSG00000183655	kelch like family member 25
KRT32	3882	ENSG00000108759	keratin 32
KRT35	3886	ENSG00000197079	keratin 35
KRT36	8689	ENSG00000126337	keratin 36
LEAP2	116842	ENSG00000164406	liver enriched antimicrobial peptide 2
LGI1	9211	ENSG00000108231	leucine rich glioma inactivated 1
LNX1	84708	ENSG00000072201	ligand of numb-protein X 1
LOC100128093	1E+08	ENSG00000232862	NA
LOC100128239	1E+08	ENSG00000204241	uncharacterized LOC100128239
LOC100128334	1E+08	ENSG00000235713	NA
LOC100129375	1E+08	ENSG00000232915	NA
LOC100129697	1E+08	ENSG00000205018	uncharacterized LOC100129697
LOC100130176	1E+08	ENSG00000180284	NA
LOC100287808	1E+08	ENSG00000277182	uncharacterized LOC100287808
LOC338963	338963	ENSG00000228141	epididymal protein pseudogene
LOC392232	392232	ENSG00000250979	NA
LOC440300	440300	ENSG00000259295	chondroitin sulfate proteoglycan 4 pseudogene
LOC642648	642648	ENSG00000235159	NA
LOC644893	644893	ENSG00000277516	NA
LOC645978	645978	ENSG00000180105	NA
LOC646700	646700	ENSG00000229060	NA
LOC648927	648927	ENSG00000232842	NA
LOC653786	653786	ENSG00000257838	otoancorin pseudogene
LOC728554	728554	ENSG00000170089	THO complex 3 pseudogene
LOC728730	728730	ENSG00000231312	uncharacterized LOC728730
LOC729654	729654	ENSG00000267056	NA
LYSMD3	116068	ENSG00000176018	LysM domain containing 3
MAGEB1	4112	ENSG00000214107	MAGE family member B1
MAGEB3	4114	ENSG00000198798	MAGE family member B3
MAGEB4	4115	ENSG00000120289	MAGE family member B4
MAGEC2	51438	ENSG00000046774	MAGE family member C2
MAN2C1	4123	ENSG00000140400	mannosidase alpha class 2C member 1
MAP2K1	5604	ENSG00000169032	mitogen-activated protein kinase kinase 1
MAP7D1	55700	ENSG00000116871	MAP7 domain containing 1
MAPK6	5597	ENSG00000069956	mitogen-activated protein kinase 6
MAPK8IP1	9479	ENSG00000121653	mitogen-activated protein kinase 8 interacting protein 1
MARCO	8685	ENSG00000019169	macrophage receptor with collagenous structure
MARK1	4139	ENSG00000116141	microtubule affinity regulating kinase 1
MARVELD2	153562	ENSG00000152939	MARVEL domain containing 2
MBL2	4153	ENSG00000165471	mannose binding lectin 2
MC4R	4160	ENSG00000166603	melanocortin 4 receptor

MCM5	4174	ENSG00000100297	minichromosome maintenance complex component 5
MIOX	55586	ENSG00000100253	myo-inositol oxygenase
MMACHC	25974	ENSG00000132763	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria
MMP11	4320	ENSG00000099953	matrix metallopeptidase 11
MOCOS	55034	ENSG00000075643	molybdenum cofactor sulfurase
MORF4L1	10933	ENSG00000185787	mortality factor 4 like 1
MPP2	4355	ENSG00000108852	membrane palmitoylated protein 2
MUC13	56667	ENSG00000173702	mucin 13, cell surface associated
MXD3	83463	ENSG00000213347	MAX dimerization protein 3
MYO15A	51168	ENSG00000091536	myosin XVA
N4BP2L1	90634	ENSG00000139597	NEDD4 binding protein 2 like 1
NADK	65220	ENSG00000008130	NAD kinase
NBPF22P	285622	ENSG00000205449	NA
NCOA6	23054	ENSG00000198646	nuclear receptor coactivator 6
NDEL1	81565	ENSG00000166579	nudE neurodevelopment protein 1 like 1
NDN	4692	ENSG00000182636	necdin, MAGE family member
NDRG2	57447	ENSG00000165795	NDRG family member 2
NDRG3	57446	ENSG00000101079	NDRG family member 3
NFIB	4781	ENSG00000147862	nuclear factor I B
NFRKB	4798	ENSG00000170322	nuclear factor related to kappaB binding protein
NKX2-4	644524	ENSG00000125816	NK2 homeobox 4
NOL6	65083	ENSG00000165271	nucleolar protein 6
NONO	4841	ENSG00000147140	non-POU domain containing, octamer-binding
NPAS3	64067	ENSG00000151322	neuronal PAS domain protein 3
NUCB1	4924	ENSG00000104805	nucleobindin 1
NUDT16L1	84309	ENSG00000168101	nudix hydrolase 16 like 1
NXF1	10482	ENSG00000162231	nuclear RNA export factor 1
OR2F2	135948	ENSG00000221910	olfactory receptor family 2 subfamily F member 2
OR52K2	119774	ENSG00000181963	olfactory receptor family 52 subfamily K member 2
OSR2	116039	ENSG00000164920	odd-skipped related transcription factor 2
P2RY4	5030	ENSG00000186912	pyrimidinergic receptor P2Y4
PABPC3	5042	ENSG00000151846	poly(A) binding protein cytoplasmic 3
PADI1	29943	ENSG00000142623	peptidyl arginine deiminase 1
PARP3	10039	ENSG00000041880	poly(ADP-ribose) polymerase family member 3
PART1	25859	ENSG00000152931	prostate androgen-regulated transcript 1 (non-protein coding)
PARVB	29780	ENSG00000188677	parvin beta
PCGF3	10336	ENSG00000185619	polycomb group ring finger 3
PCSK1N	27344	ENSG00000102109	proprotein convertase subtilisin/kexin type 1 inhibitor
PCSK6	5046	ENSG00000140479	proprotein convertase subtilisin/kexin type 6
PDE6H	5149	ENSG00000139053	phosphodiesterase 6H
PDHB	5162	ENSG00000168291	pyruvate dehydrogenase (lipoamide) beta
PDPK1	5170	ENSG00000140992	3-phosphoinositide dependent protein kinase 1
PDSS2	57107	ENSG00000164494	decaprenyl diphosphate synthase subunit 2
PGM2	55276	ENSG00000169299	phosphoglucomutase 2

PHKG2	5261	ENSG00000156873	phosphorylase kinase catalytic subunit gamma 2
PIP4K2C	79837	ENSG00000166908	phosphatidylinositol-5-phosphate 4-kinase type 2 gamma
PITPNM3	83394	ENSG00000091622	PITPNM family member 3
PLD4	122618	ENSG00000166428	phospholipase D family member 4
PNPLA5	150379	ENSG00000100341	patatin like phospholipase domain containing 5
POLDIP2	26073	ENSG00000004142	DNA polymerase delta interacting protein 2
PPCDC	60490	ENSG00000138621	phosphopantothenoylcysteine decarboxylase
PPEF2	5470	ENSG00000156194	protein phosphatase with EF-hand domain 2
PPL	5493	ENSG00000118898	periplakin
PPP1CA	5499	ENSG00000172531	protein phosphatase 1 catalytic subunit alpha
PPP1R2P3	153743	ENSG00000231989	protein phosphatase 1 regulatory inhibitor subunit 2 pseudogene 3
PRDM6	93166	ENSG00000061455	PR/SET domain 6
PRELID1	27166	ENSG00000169230	PRELI domain containing 1
PRKAR1B	5575	ENSG00000188191	protein kinase cAMP-dependent type I regulatory subunit beta
PRMT7	54496	ENSG00000132600	protein arginine methyltransferase 7
PROM1	8842	ENSG00000007062	prominin 1
PRR16	51334	ENSG00000184838	proline rich 16
PSMC3IP	29893	ENSG00000131470	PSMC3 interacting protein
PTDSS1	9791	ENSG00000156471	phosphatidylserine synthase 1
PTGES3	10728	ENSG00000110958	prostaglandin E synthase 3
PTK6	5753	ENSG00000101213	protein tyrosine kinase 6
PTPRN	5798	ENSG00000054356	protein tyrosine phosphatase, receptor type N
PUF60	22827	ENSG00000179950	poly(U) binding splicing factor 60
PXMP4	11264	ENSG00000101417	peroxisomal membrane protein 4
RAB24	53917	ENSG00000169228	RAB24, member RAS oncogene family
RAB41	347517	ENSG00000147127	RAB41, member RAS oncogene family
RABL3	285282	ENSG00000144840	RAB, member of RAS oncogene family like 3
RASGRF2	5924	ENSG00000113319	Ras protein specific guanine nucleotide releasing factor 2
RASL11A	387496	ENSG00000122035	RAS like family 11 member A
RCVRN	5957	ENSG00000109047	recoverin
REEP2	51308	ENSG00000132563	receptor accessory protein 2
RFC5	5985	ENSG00000111445	replication factor C subunit 5
RGL4	266747	ENSG00000159496	ral guanine nucleotide dissociation stimulator like 4
RIPPLY2	134701	ENSG00000203877	ripply transcriptional repressor 2
RNF25	64320	ENSG00000163481	ring finger protein 25
RNF34	80196	ENSG00000170633	ring finger protein 34
RNF5	6048	ENSG00000204308	ring finger protein 5
ROBO1	6091	ENSG00000169855	roundabout guidance receptor 1
ROCK1	6093	ENSG00000067900	Rho associated coiled-coil containing protein kinase 1
RPL10A	4736	ENSG00000198755	ribosomal protein L10a
RPL21	6144	ENSG00000122026	ribosomal protein L21
RPL21P28	1E+08	ENSG00000220749	ribosomal protein L21 pseudogene 28
RPL21P44	402176	ENSG00000229585	NA
RTEL1	51750	ENSG00000026036	regulator of telomere elongation helicase 1

RTN4IP1	84816	ENSG00000130347	reticulon 4 interacting protein 1
RUVBL2	10856	ENSG00000183207	RuvB like AAA ATPase 2
SARM1	23098	ENSG00000004139	sterile alpha and TIR motif containing 1
SCHIP1	29970	ENSG00000151967	schwannomin interacting protein 1
SEC14L5	9717	ENSG00000103184	SEC14 like lipid binding 5
SETD2	29072	ENSG00000181555	SET domain containing 2
SETD4	54093	ENSG00000185917	SET domain containing 4
SGSM1	129049	ENSG00000167037	small G protein signaling modulator 1
SH2D7	646892	ENSG00000183476	SH2 domain containing 7
SH3BP2	6452	ENSG00000087266	SH3 domain binding protein 2
SIGLEC12	89858	ENSG00000254521	sialic acid binding Ig like lectin 12 (gene/pseudogene)
SIGLEC14	1E+08	ENSG00000254415	sialic acid binding Ig like lectin 14
SIGLEC15	284266	ENSG00000197046	sialic acid binding Ig like lectin 15
SIGLEC5	8778	ENSG00000105501	sialic acid binding Ig like lectin 5
SIGLEC7	27036	ENSG00000168995	sialic acid binding Ig like lectin 7
SIGLEC8	27181	ENSG00000105366	sialic acid binding Ig like lectin 8
SIGLEC9	27180	ENSG00000129450	sialic acid binding Ig like lectin 9
SIRT6	51548	ENSG00000077463	sirtuin 6
SLAIN2	57606	ENSG00000109171	SLAIN motif family member 2
SLC19A1	6573	ENSG00000173638	solute carrier family 19 member 1
SLC24A4	123041	ENSG00000140090	solute carrier family 24 member 4
SLC25A1	6576	ENSG00000100075	solute carrier family 25 member 1
SLC25A37	51312	ENSG00000147454	solute carrier family 25 member 37
SLC27A5	10998	ENSG00000083807	solute carrier family 27 member 5
SLC29A2	3177	ENSG00000174669	solute carrier family 29 member 2
SLC38A4	55089	ENSG00000139209	solute carrier family 38 member 4
SLC47A1	55244	ENSG00000142494	solute carrier family 47 member 1
SLIT2	9353	ENSG00000145147	slit guidance ligand 2
SMARCAD1	56916	ENSG00000163104	SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1
SMARCB1	6598	ENSG00000099956	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily b, member 1
SNAPC2	6618	ENSG00000104976	small nuclear RNA activating complex polypeptide 2
SNORA27	619499	ENSG00000207051	small nucleolar RNA, H/ACA box 27
SNORA59B	677882	ENSG00000266079	small nucleolar RNA, H/ACA box 59B
SNORD102	26771	ENSG00000207500	small nucleolar RNA, C/D box 102
SNORD126	1E+08	ENSG00000238344	NA
SOHLH1	402381	ENSG00000165643	spermatogenesis and oogenesis specific basic helix-loop-helix 1
SPIB	6689	ENSG00000269404	Spi-B transcription factor
SPNS2	124976	ENSG00000183018	sphingolipid transporter 2
SPTY2D1	144108	ENSG00000179119	SPT2 chromatin protein domain containing 1
SRD5A2	6716	ENSG00000277893	steroid 5 alpha-reductase 2
SSFA2	6744	ENSG00000138434	sperm specific antigen 2
SSTR1	6751	ENSG00000139874	somatostatin receptor 1
SSX8	280659	ENSG00000157965	NA

ST8SIA2	8128	ENSG00000140557	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 2
STK11IP	114790	ENSG00000144589	serine/threonine kinase 11 interacting protein
STMN3	50861	ENSG00000197457	stathmin 3
STX1B	112755	ENSG00000099365	syntaxin 1B
STXBP4	252983	ENSG00000166263	syntaxin binding protein 4
SUOX	6821	ENSG00000139531	sulfite oxidase
SWAP70	23075	ENSG00000133789	SWAP switching B-cell complex subunit 70
SYNC	81493	ENSG00000162520	syncoilin, intermediate filament protein
SYTL1	84958	ENSG00000142765	synaptotagmin like 1
TAPT1	202018	ENSG00000169762	transmembrane anterior posterior transformation 1
TAX1BP3	30851	ENSG00000213977	Tax1 binding protein 3
TBC1D10C	374403	ENSG00000175463	TBC1 domain family member 10C
TBC1D2B	23102	ENSG00000167202	TBC1 domain family member 2B
TDRD1	56165	ENSG00000095627	tudor domain containing 1
TESK2	10420	ENSG00000070759	testis-specific kinase 2
THBD	7056	ENSG00000178726	thrombomodulin
THOC2	57187	ENSG00000125676	THO complex 2
THOC3	84321	ENSG00000051596	THO complex 3
TMEM104	54868	ENSG00000109066	transmembrane protein 104
TMEM189	387521	ENSG00000240849	transmembrane protein 189
TMEM199	147007	ENSG00000244045	transmembrane protein 199
TMEM81	388730	ENSG00000174529	transmembrane protein 81
TNFAIP8	25816	ENSG00000145779	TNF alpha induced protein 8
TNNT3	7140	ENSG00000130595	troponin T3, fast skeletal type
TOR1AIP1	26092	ENSG00000143337	torsin 1A interacting protein 1
TOR1AIP2	163590	ENSG00000169905	torsin 1A interacting protein 2
TP53	7157	ENSG00000141510	tumor protein p53
TPCN1	53373	ENSG00000186815	two pore segment channel 1
TPM3	7170	ENSG00000143549	tropomyosin 3
TPM4	7171	ENSG00000167460	tropomyosin 4
TPPP2	122664	ENSG00000179636	tubulin polymerization promoting protein family member 2
TPPP3	51673	ENSG00000159713	tubulin polymerization promoting protein family member 3
TRAM2	9697	ENSG00000065308	translocation associated membrane protein 2
TRAPPC3	27095	ENSG00000054116	trafficking protein particle complex 3
TREML2	79865	ENSG00000112195	triggering receptor expressed on myeloid cells like 2
TRIB2	28951	ENSG00000071575	tribbles pseudokinase 2
TRIM60	166655	ENSG00000176979	tripartite motif containing 60
TSPAN14	81619	ENSG00000108219	tetraspanin 14
TTC39C	125488	ENSG00000168234	tetratricopeptide repeat domain 39C
TTTY14	83869	ENSG00000176728	NA
TYMS	7298	ENSG00000176890	thymidylate synthetase
UBE2B	7320	ENSG00000119048	ubiquitin conjugating enzyme E2 B
UBE2Z	65264	ENSG00000159202	ubiquitin conjugating enzyme E2 Z
UBXN6	80700	ENSG00000167671	UBX domain protein 6

UGT3A2	167127	ENSG00000168671	UDP glycosyltransferase family 3 member A2
UPK1A	11045	ENSG00000105668	uroplakin 1A
UPK3B	1.05E+08	ENSG00000243566	uroplakin 3B
UQCRQ	27089	ENSG00000164405	ubiquinol-cytochrome c reductase complex III subunit VII
VAC14	55697	ENSG00000103043	Vac14, PIKFYVE complex component
VGLL4	9686	ENSG00000144560	vestigial like family member 4
VILL	50853	ENSG00000136059	villin like
VPREB3	29802	ENSG00000128218	V-set pre-B cell surrogate light chain 3
VSIG8	391123	ENSG00000243284	V-set and immunoglobulin domain containing 8
VTN	7448	ENSG00000109072	vitronectin
VTRNA1-1	56664	ENSG00000199990	vault RNA 1-1
VTRNA1-2	56663	ENSG00000202111	vault RNA 1-2
VTRNA1-3	56662	ENSG00000202515	vault RNA 1-3
VWA1	64856	ENSG00000179403	von Willebrand factor A domain containing 1
WDR13	64743	ENSG00000101940	WD repeat domain 13
WDR6	11180	ENSG00000178252	WD repeat domain 6
WDR61	80349	ENSG00000140395	WD repeat domain 61
WDR73	84942	ENSG00000177082	WD repeat domain 73
WISP2	8839	ENSG00000064205	WNT1 inducible signaling pathway protein 2
ZBTB40	9923	ENSG00000184677	zinc finger and BTB domain containing 40
ZBTB45	84878	ENSG00000119574	zinc finger and BTB domain containing 45
ZCCHC2	54877	ENSG00000141664	zinc finger CCHC-type containing 2
ZFYVE28	57732	ENSG00000159733	zinc finger FYVE-type containing 28
ZMAT2	153527	ENSG00000146007	zinc finger matrin-type 2
ZNF135	7694	ENSG00000176293	zinc finger protein 135
ZNF180	7733	ENSG00000167384	zinc finger protein 180
ZNF219	51222	ENSG00000165804	zinc finger protein 219
ZNF341	84905	ENSG00000131061	zinc finger protein 341
ZNF35	7584	ENSG00000169981	zinc finger protein 35
ZNF423	23090	ENSG00000102935	zinc finger protein 423
ZNF501	115560	ENSG00000186446	zinc finger protein 501
ZNF674	641339	ENSG00000251192	zinc finger protein 674
ZXDA	7789	ENSG00000198205	zinc finger, X-linked, duplicated A
ZXDB	158586	ENSG00000198455	zinc finger, X-linked, duplicated B

APPENDIX 4.2 Metrics of RNA-seq samples from NR2F2 siRNA in ALT+ cell lines.

Sample	Mapping				Counting of uniquely mapped PE reads			
	#PE Reads	Uniqu e (%)	Multipl e (%)	Unmappe d (%)	#Reads	Genes (%)	No Features (%)	Ambiguit y (%)
WV_CNT1	37352573	89.18	1.59	9.22	33312599	85.31	7.84	6.85
WV_CNT2	38759150	86.53	1.62	11.85	33538376	85.54	7.44	7.02
WV_CNT3	39988646	87.91	1.61	10.48	35153214	85.99	7.01	7.00
WV_KD1	39098450	88.95	1.53	9.52	34778296	85.24	7.63	7.13
WV_KD2	37302647	86.66	1.51	11.83	32327763	85.17	7.66	7.16
WV_KD3	39670887	88.81	1.49	9.70	35231872	85.13	7.82	7.05
WI38VA13_2RA_CN T1	35478095	85.72	1.46	12.82	30413039	86.45	6.66	6.89
WI38VA13_2RA_CN T2	39013543	85.97	1.44	12.59	33540752	86.14	7.25	6.62
WI38VA13_2RA_CN T3	34352727	82.44	1.39	16.18	28319095	86.24	7.05	6.70
WI38VA13_2RA_KD 1	36584075	85.90	1.40	12.70	31426756	85.45	7.85	6.70
WI38VA13_2RA_KD 2	41013437	86.80	1.41	11.79	35599370	85.07	8.36	6.56
WI38VA13_2RA_KD 3	34207461	88.71	1.45	9.84	30347057	85.20	8.18	6.61
U2OS_CNT1	35618402	85.57	2.11	12.33	30477596	83.24	10.31	6.45
U2OS_CNT2	32779420	85.42	2.09	12.49	27999858	82.74	10.94	6.32
U2OS_CNT3	46517942	83.45	2.20	14.36	38817771	83.23	10.08	6.69
U2OS_KD1	31090137	86.04	2.22	11.74	26750809	80.43	13.58	5.99
U2OS_KD2	31618459	86.67	2.35	10.98	27402325	79.86	14.35	5.80
U2OS_KD3	30425192	85.94	2.22	11.83	26147924	81.24	12.66	6.10

Metrics of RNA-seq samples from NR2F2 siRNA in ALT+ cell lines.

APPENDIX 4.3 Scatter plots of gene expression profiles of RNA-seq samples from siNR2F2 in ALT+ cell lines
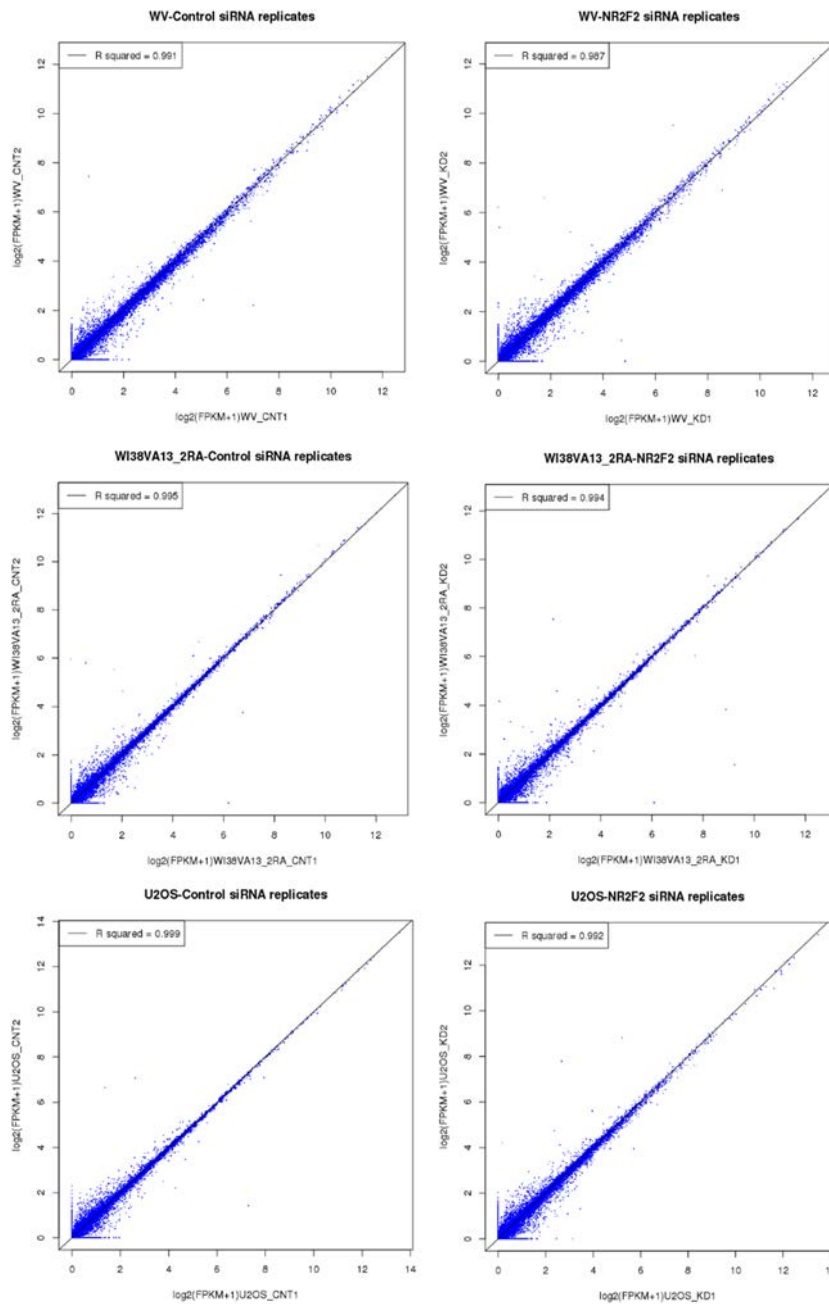
Scatter plots of gene expression profiles of RNA-seq samples from siNR2F2 in ALT+ cell lines. The scatter plots diagrams demonstrate the correlation coefficient between samples. Data points represent log2(FPKM+1). Data points represent log2(FPKM+1) of expressed genes. All data points are arrayed clearly along the diagonal lines (black). R^2, the square of the Pearson coefficient.

APPENDIX 4.4 Heatmap of the correlation coefficient between RNA samples are shown from siNR2F2 in ALT+ cell lines
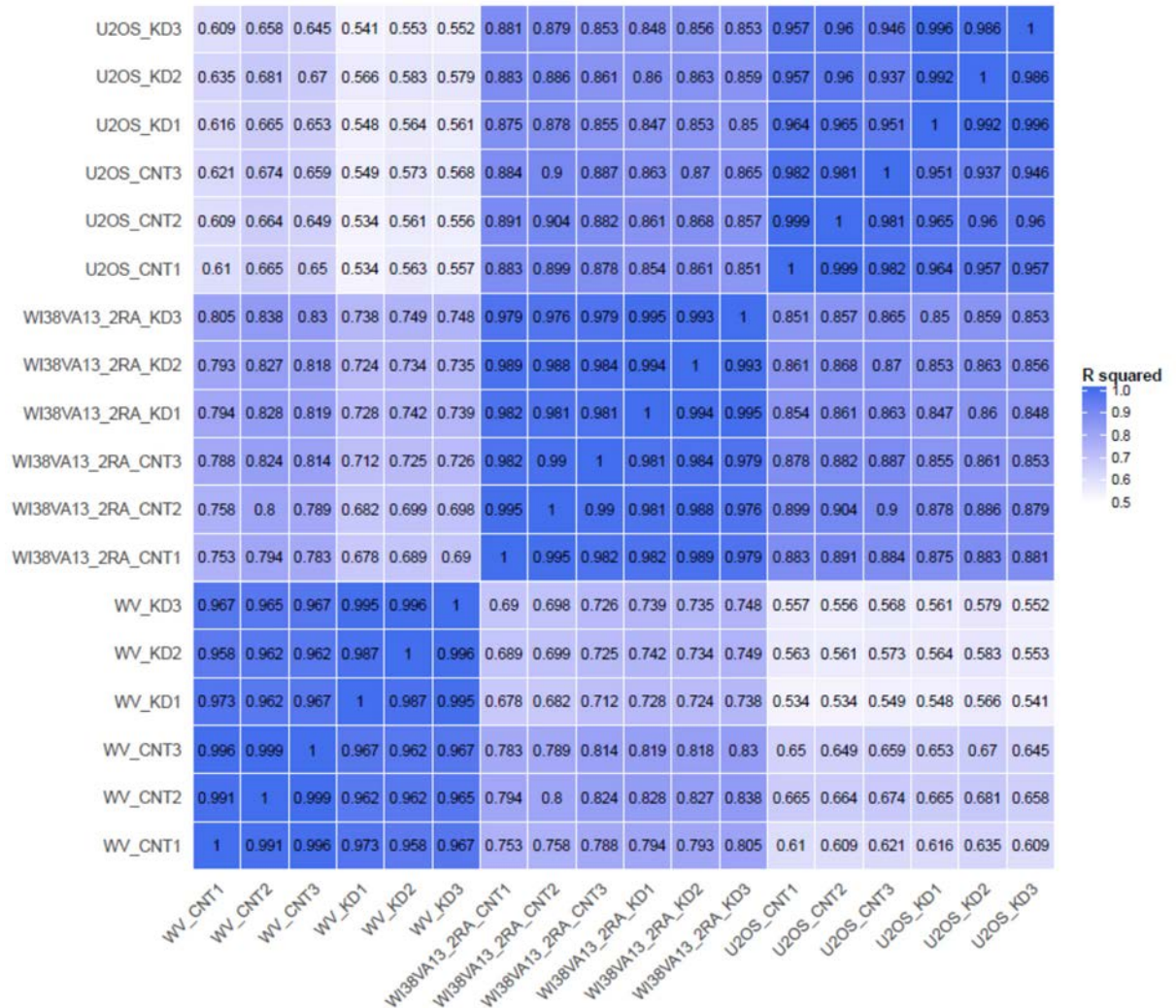
Heatmap of the correlation coefficient between RNA samples are shown from siNR2F2 in ALT+ cell lines. R^2, the square of the Pearson coefficient. The closer R^2 is to 1, the greater the similarity of the samples.

APPENDIX 4.5 GO enrichment of DE genes upon NR2F2 downregulation in WV cells

GO	ID	Description	p.adjust	GeneID	Count
BP	GO:0043062	extracellular structure organization	1.21E-05	ABI3BP/ITGA10/TGFBR1/LRP1/FURIN/CCDC80/ITGA2/NID2/OLFML2A/SOX9/MMP8/COL7A1/SH3PXD2B/TGFBI/COL16A1/MMP13/DDR1/ITGA1/MMP2/ITGA11/TNC/ITGAX/MMP1/CTGF/LOX/SPP1/FBN1/ITGB3/POSTN/ADAMTS5/LAMA1/ITGB2/ITGA7/VIT/COL9A2/VCAN/CTSK/BCL3/PECAM1/MFAP4/MMP3/DPP4	42
BP	GO:0030335	positive regulation of cell migration	9.42E-05	MYADM/HBEGF/TGFBR1/PIK3CD/ITGA2/SOX9/SDCBP/NRP2/LPAR1/HDAC9/GLIPR2/MDM2/IQGAP1/CARMIL1/ROCK2/DOCK7/PDGFRB/CDH13/SMURF2/SOD2/PDGFRA/RELN/GPNMB/CXCL8/EPB41L5/CORO1A/IL1R1/ITGB3/PREX1/HSPB1/POSTN/ONECUT2/CPEB1/F3/PTGS2/ZNF703/PDGFD/CSF1R/IL1A/NTRK3/CCL20/CSF1/ATP8A1/HMCN2/EPHA1/IL23A	46
BP	GO:0048514	blood vessel morphogenesis	1.57E-04	NR2F2/SHC1/SCG2/TGFBR1/LRP1/GNA13/ELK3/CLIC4/EPHB2/NRP2/TGFBI/ADGRB2/NFATC4/XDH/VASH2/HDAC9/RBPJ/MMP2/RORA/ROCK2/PDGFRB/C3/CDH13/EIF2AK3/CTGF/SIX1/LOX/E2F8/HEY1/PDGFRA/PIK3CB/STAT1/GPNMB/CXCL8/ITGB3/TNFAIP3/HSPB1/NPPB/F3/PTGS2/ITGB2/ZFPM2/JAG1/DLL4/IL1A/HES1/PLCD1/WNT7B/PIK3CG/NOTCH4/STRA6/EPHA1	52
BP	GO:0001101	response to acid chemical	7.82-04	ITGA2/GSN/SOX9/CCNB1/GCLM/COL16A1/NFATC4/MMP2/PDGFRB/ATF2/BRINP3/TNC/EGR1/CTGF/SIX1/ID3/CPEB4/SESN1/TH/CPEB1/GIPR/PTGS2/ASNS/UCP2/TRIML2/HSF1/ABCC2/OSR1/PDGFD/CPEB3/CDO1/NTRK3/CYP2E1/WNT7B/GNRH1/FABP3/HOXA2	37
BP	GO:0007565	female pregnancy	2.08E-03	NR2F2/ITGA2/STC1/IGFBP5/STAT5B/COL16A1/DDR1/APOL2/MMP2/PAPPA/SMURF2/ARHGDIB/SPP1/FOS/PSG4/PTGS2/UCP2/CRHBP/HSF1/ABCC2/PSG2/ACVR1C/OXTR/PRLR/GNRH1	25
BP	GO:0044706	multi-multicellular organism process	2.64E-03	NR2F2/ITGA2/STC1/SH3PXD2B/IGFBP5/STAT5B/COL16A1/DDR1/APOL2/MMP2/PAPPA/SMURF2/ARHGDIB/SPP1/FOS/PSG4/PTGS2/UCP2/CRHBP/HSF1/ABCC2/PSG2/ACVR1C/OXTR/PRLR/GNRH1/DDO	27
BP	GO:0001655	urogenital system development	2.64E-03	TGFBR1/ADAMTS1/SMAD5/SOX9/RRM2B/EPHB2/ACTA2/IQGAP1/NKX3-1/PDGFRB/TP73/ODC1/FRS2/TNC/EGR1/SIX1/PDGFRA/GDNF/STAT1/ID3/FBN1/KIRREL3/SALL1/CD24/LGR5/OSR1/PDGFD/JAG1/WNT2B/PECAM1/HES1/ROBO2/WNT7B/STRA6/WNK4	35
BP	GO:0001822	kidney development	2.89E-03	TGFBR1/ADAMTS1/SMAD5/SOX9/RRM2B/ACTA2/IQGAP1/NKX3-1/PDGFRB/TP73/ODC1/EGR1/SIX1/PDGFRA/GDNF/STAT1/ID3/FBN1/KIRREL3/SALL1/CD24/OSR1/PDGFD/JAG1/WNT2B/PECAM1/HES1/ROBO2/WNT7B/STRA6/WNK4	31
BP	GO:0001503	ossification	2.92E-03	TMEM64/STC1/SEMA7A/SMAD5/SOX9/MRC2/SH3PXD2B/IGFBP5/CTHRC1/IL6ST/MMP13/CHSY1/RBPJ/KREMEN1/MMP2/ITGA11/ISG15/EIF2AK3/TNC/CTGF/SPP1/GPNMB/CHRDL1/ID3/AREG/RASSF2/PEX7/PTGS2/PTH1R/GDPD2/DHRS3/OSR1/JAG1/VCAN/CTSK/CSF1/WNT7B/PKDCC/HOXA2	39
BP	GO:0031589	cell-substrate adhesion	5.05E-03	ABI3BP/MYADM/ITGA10/CCDC80/ITGA2/NID2/GBP1/BVES/COL16A1/DDR1/ITGA1/IQGAP1/CARMIL1/ITGA11/ROCK2/CDH13/CTGF/PIK3CB/EPB41L5/SGCE/CORO1A/ITGB3/PREX1/DISC1/POSTN/ONECUT2/ITGB2/SORBS1/ITGA7/MERTK/VIT/CSF1/EPHA1	33

BP	GO:003 5239	tube morphogenesis	5.05E-03	GNA13/MICAL2/PLXNB2/DACT1/CLIC4/SOX9/IP MK/CTHRC1/NFATC4/DDR1/TULP3/RBPJ/MIB1/N KX3-1/FZD6/TNC/SIX1/PDGFRA/GDNF/EPB41L5/AREG/ SALL1/SALL4/LGR5/OSR1/CSF1R/DLL4/WNT2B/H ES1/CSF1/CELSR1/BBS5/TGM2/NOTCH4/STRA6/W NK4	36
BP	GO:004 8771	tissue remodeling	5.58E-03	TMEM64/INPP5D/IGFBP5/CTHRC1/RBPJ/MDM2/S PP1/GPNMB/ITGB3/RASSF2/TNFAIP3/PTH1R/MER TK/CSF1R/JAG1/DLL4/CTSK/IL1A/TGM2/IL23A	20
BP	GO:007 1900	regulation of protein serine/threonine kinase activity	8.41E-03	NR2F2/SHC1/CDKN1A/PDCD4/TGFBR1/CDC25A/E IF2AK2/CKS2/CCNB1/FBXO7/DUSP7/LPAR1/CAB3 9/ERN1/TNFRSF10B/ITGA1/MAP3K12/IQGAP1/DU SP10/PDGFRB/FAM212B/TRIB2/TP73/ALK/FRS2/N GF/PTPN22/RGS3/PIK3CB/CDK5R2/TNFAIP3/HIPK 3/HSPB1/CD24/SFN/CDKN2B/PKMYT1/IRAK2/PD GFD/CSF1R/DUSP8/NTRK3/WNT7B/PIK3CG	44
BP	GO:001 4909	smooth muscle cell migration	9.41E-03	LRP1/ITGA2/IGFBP5/DDR1/LPAR1/MDM2/IQGAP1 /DOCK7/PDGFRB/ITGB3/POSTN/PDGFD	12
BP	GO:200 0134	negative regulation of G1/S transition of mitotic cell cycle	9.41E-03	CDKN1A/DACT1/AURKA/CCNB1/FBXO7/BTG2/M DM2/TFDP2/TP73/RPS27L/GPNMB/FHL1/ZNF385A /SFN/CDKN2B	15
BP	GO:009 7193	intrinsic apoptotic signaling pathway	9.41E-03	CDKN1A/TMEM109/CHAC1/BBC3/E2F2/PHLDA3/ RRM2B/NFATC4/IFI16/ERN1/TNFRSF10B/MDM2/I TPR1/NKX3-1/TP73/BDKRB2/EIF2AK3/RPS27L/SOD2/NUPR1/Z MAT3/DDIT3/HSPB1/ZNF385A/CD24/SFN/PTGS2/B CL3/BMF/HSPA1A	30
BP	GO:005 1961	negative regulation of nervous system development	1.02E-02	LRP1/GPR173/STX1B/SOX9/EPHB2/DPYSL3/NFAT C4/LPAR1/TULP3/DCC/MDM2/DUSP10/MIB1/TP73 /EIF2AK3/HOOK3/SPP1/JAG1/TNFRSF8/NTRK3/HE S1/ISL2/ROBO2/GNRH1/HOXA2/ARHGAP4/PAX6/ TLX2	28
BP	GO:003 3002	muscle cell proliferation	1.02E-02	HBEGF/TGFBR1/ITGA2/SKP2/CCNB1/IGFBP5/ERN 1/RBPJ/MDM2/MMP2/PDGFRB/CDH13/TP73/SIX1/ STAT1/KCNK2/TNFAIP3/PTGS2/ZFPM2/PDGFD/T GM2	21
BP	GO:005 1607	defense response to virus	1.22E-02	SAMHD1/TRIM22/EIF2AK2/PUM2/GBP1/MX2/IFIT 1/IFI16/RIOK3/STAT2/LYST/ISG15/IRF9/DNAJC3/S TAT1/PLSCR1/C19orf66/TNFAIP3/ISG20/IFIH1/IFI4 4L/IL23A/IFNE/DHX58/IFITM3/OAS3	26
BP	GO:009 0130	tissue migration	1.53E-02	NR2F2/HBEGF/SCG2/TGFBR1/ITGA2/STC1/SOX9/ NRP2/HDAC9/GLIPR2/ACTA2/ROCK2/S100A2/CD H13/PLEKHG5/EPB41L5/ITGB3/EPPK1/HSPB1/PTG S2/ITGB2/MCC/DLL4/DPP4/ACTC1/PTPRR	26
CC	GO:000 5578	proteinaceous extracellular matrix	1.47E-03	ABI3BP/LTBP2/ADAMTS1/CCDC80/NID2/ADAMT SL1/OLFML2A/MMP8/COL7A1/CTHRC1/TGFBI/C OL16A1/MMP13/MEGF9/MMP2/P3H2/COL24A1/TN C/MMP1/CTGF/LOX/RELN/FBN1/ADAMTS7/POST N/ADAMTS5/LAMA1/VIT/COL9A2/VCAN/WNT2B/ CPA6/CHAD/SERPINA1/MFAP4/WNT7B/MMP3/M AMDC2/HMCN2	39
CC	GO:000 8305	integrin complex	3.03E-02	ITGA10/ITGA2/ITGA1/ITGA11/ITGAX/ITGB3/ITGB 2/ITGA7	8
CC	GO:009 8636	protein complex involved in cell adhesion	3.97E-02	ITGA10/ITGA2/ITGA1/ITGA11/ITGAX/ITGB3/ITGB 2/ITGA7	8
CC	GO:004 5121	membrane raft	4.15E-02	MYADM/IRS1/CAV2/TGFBR1/FURIN/INPP5D/CLIP 3/BVES/SDCBP/RHOQ/ITGA1/FAS/IQGAP1/CDH13 /SMURF2/PPP2R1B/PLSCR1/CAVIN2/CD24/PTGS2/ SORBS1/GRIP1/PECAM1/PDZK1/PLLP/DPP4/NTSR 1/TRPM8	28

CC	GO:0019005	SCF ubiquitin ligase complex	4.43E-02	SKP2/CKS2/SPOPL/BTBD1/FBXO7/CCNF/FBXO32/FBXW8/BTBD3	9
CC	GO:0030424	axon	4.43E-02	BSN/SYT11/NCS1/ITGA2/EPHB2/AURKA/SYP/CRYAB/DCC/HOMER1/MAP3K12/IQGAP1/DOCK7/IGF2BP1/STAT1/KCNK2/CORO1A/KCNH1/KIRREL3/SLC17A7/DISC1/TH/HSPB1/CPT1C/PALLD/AZIN2/CRHBP/ERMN/BOC/ADGRL3/ROBO2/NTSR1/ANK1/C4B	34
CC	GO:0005884	actin filament	4.43E-02	CD2AP/PLS3/DPYSL3/RHOQ/IQGAP1/CARMIL1/CORO1A/MYO5A/AIF1L/PALLD/GDPD2/ACTC1	12
MF	GO:0019199	transmembrane receptor protein kinase activity	7.80E-04	CRIM1/NPTN/TGFBR1/EPHB2/NRP2/DDR1/PDGFRB/ALK/PDGFRA/ROS1/EPHB6/MERTK/ACVR1C/CSF1R/NTRK3/EPHA1	16
MF	GO:0004714	transmembrane receptor protein tyrosine kinase activity	7.80E-04	CRIM1/NPTN/EPHB2/NRP2/DDR1/PDGFRB/ALK/PDGFRA/ROS1/EPHB6/MERTK/CSF1R/NTRK3/EPHA1	14
MF	GO:0005161	platelet-derived growth factor receptor binding	8.47E-03	ERN1/PDGFRB/PDGFRA/IL1R1/ITGB3/PDGFD	6
MF	GO:0019838	growth factor binding	8.47E-03	SHC1/CRIM1/LTBP2/TGFBR1/FURIN/NRP2/IGFBP4/IGFBP5/IL6ST/PDGFRB/CTGF/PDGFRA/LIFR/IL1R1/ITGB3/ACVR1C/TNFRSF8/NTRK3	18
MF	GO:0001078	transcriptional repressor activity, RNA polymerase II core promoter proximal region sequence-specific binding	8.47E-03	BTG2/NFATC4/IFI16/BACH1/E2F8/MXD1/NFATC2/SALL1/ATF3/HMGA2/ZFPM2/HSF1/BHLHE41/KLF17/HES1/HOXA2/FOXQ1	17
MF	GO:0005518	collagen binding	8.47E-03	ABI3BP/ITGA10/ITGA2/NID2/MRC2/ADGRG6/TGFBI/MMP13/DDR1/ITGA1/ITGA11/CTSK	12
MF	GO:0001227	transcriptional repressor activity, RNA polymerase II transcription regulatory region sequence-specific binding	1.96E-02	BTG2/NFATC4/IFI16/BACH1/E2F8/HIVEP1/MXD1/NFATC2/SALL1/ATF3/HMGA2/KLF12/ZFPM2/HSF1/BHLHE41/PCBP3/KLF17/HES1/MSC/HOXA2/FOXQ1	21
MF	GO:0070851	growth factor receptor binding	1.96E-02	SHC1/HBEGF/CD2AP/FGF5/NPTN/SDCBP/IL6ST/ERN1/PDGFRB/FRS2/PDGFRA/PLSCR1/AREG/IL1R1/ITGB3/PDGFD/IL1A	17
MF	GO:0004713	protein tyrosine kinase activity	4.49E-02	HBEGF/CRIM1/FGF5/NPTN/EIF2AK2/WEE1/EPHB2/NRP2/STAT5B/DDR1/PDGFRB/ALK/PDGFRA/ROS1/EPHB6/MERTK/CSF1R/NTRK3/PKDCC/EPHA1	20

GO enrichment of DE genes upon NR2F2 downregulation in WV cells. Differentially expressed genes in the WV cell line (fold change >1.6, FDR <0.05) were mapped to GO terms using David platform, where ALL GO terms were tested. The enrichment results

was summarised by allowed medium similarity=0.7 using REViGO online tool (Supek *et al.*, 2011). GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.

APPENDIX 4.6 GO enrichment of DE genes upon NR2F2 downregulation in WI38VA13_2RA cells

GO	ID	Description	p.adjust	geneID	Count
BP	GO:00 30335	positive regulation of cell migration	3.47E-08	CXCL1/SOD2/NRP2/PIK3CD/TGFBR1/CCBE1/GAS6/MYADM/HAS2/WNT5B/SEMA5A/CDH13/LYN/CAPN7/CSF1/ANXA3/ROCK2/TIAM1/CXCL8/VEGFC/HSPB1/TGFB1/PTGS2/ADGRA2/CORO1A/POSTN/CCL3/MMP9/F2R/CD274/CCL5/IRS2/CCL20/LRRC15/RET/CPEB1/GCNT2/IL6/IL1R1/FOXF1/CCR1/PDGFD/TEK/CSF1R/EPB41L4B/IL23A/HMCN2/TNFAIP6	48
BP	GO:00 48514	blood vessel morphogenesis	2.13E-07	SHC1/NRP2/SLIT2/EPAS1/NR2F2/TGFBR1/MMP2/CCBE1/PIK3CB/SEMA3E/HAS2/SEMA5A/CDH13/ELK3/ANXA3/NOV/LRP1/TGFBR3/IL18/ROCK2/CXCL8/VEGFC/HSPB1/SULF1/TGFB1/SIX1/PTGS2/ADGRA2/ITGB2/PLXDC1/EIF2AK3/PTGIS/ADGRB2/NFATC4/COL8A1/FZD8/EGF/PLCD1/EPHB3/IL6/HOXA1/NOTCH3/FOXF1/WNT7B/IL1B/STRA6/ACVRL1/C3/TEK/EDNRA/ESM1/NOTCH4/COL15A1	53
BP	GO:00 60337	type I interferon signaling pathway	1.57E-05	ISG15/MX2/RSAD2/IFIT1/IFI6/OASL/OAS2/ISG20/IRF9/IFIT2/IFI27/IFITM1/OAS1/USP18/GBP2/IRF7/OAS3	17
BP	GO:00 01501	skeletal system development	7.54E-05	IGFBP4/FAM20C/TGFBR1/MMP2/ADAMTS12/HMGA2/SH3PXD2B/HAS2/WNT5B/TULP3/RASSF2/NOV/DCHS1/SULF1/TGFB1/SIX1/VCAN/CHSY1/EIF2AK3/SULF2/CHST11/POSTN/CDH11/PKDCC/MMP9/ZFPM1/WNT10B/HOXC10/STC1/HOXD10/ZEB1/EN1/PTGER4/HOXA1/FRZB/AXIN2/NKX3-2/WNT7B/ALX4/CDKN1C/ACVRL1/TNFRSF11B/TEK/SIGLEC15/EFEMP1/DLL3	46
BP	GO:00 02062	chondrocyte differentiation	1.59E-04	TGFBR1/ADAMTS12/HMGA2/WNT5B/NOV/SULF1/TGFB1/CHSY1/EIF2AK3/SULF2/CHST11/PKDCC/WNT10B/AXIN2/NKX3-2/ACVRL1/EFEMP1	17
BP	GO:00 06939	smooth muscle contraction	1.72E-04	GDNF/ROCK2/ADRB2/BDKRB2/SULF1/ADA/PTGS2/SULF2/ATP1A2/F2R/EDNRB/ACTA2/ABAT/HTR1D/EDNRA/HMCN2/CHRNB2	17
BP	GO:00 01822	kidney development	2.30E-04	GDNF/SLIT2/TGFBR1/RRM2B/ALDH1A2/HAS2/CTNNBIP1/MMP17/DCHS1/AMPD2/SULF1/TGFB1/SIX1/FOXD1/SULF2/LIF/MME/RET/ACTA2/NOTCH3/WNT7B/STRA6/PDGFD/CDKN1C/EPHA7/NID1/KLF15/TEK/KIF26B/WNT6	30
BP	GO:00 61448	connective tissue development	5.21E-04	TGFBR1/ADAMTS12/HMGA2/SH3PXD2B/WNT5B/CSF1/NOV/SULF1/TGFB1/FOXD1/CHSY1/EIF2AK3/SULF2/CHST11/PKDCC/WNT10B/ACTA2/STC1/ZEB1/FRZB/AXIN2/NKX3-2/WNT7B/PDGFD/ACVRL1/EFEMP1/PPARGC1A	27
BP	GO:00 90130	tissue migration	5.58E-04	NRP2/SLIT2/DPP4/NR2F2/TGFBR1/CCBE1/HAS2/SEMA5A/CDH13/CAPN7/ANXA3/NOV/ROCK2/VEGFC/HSPB1/TGFB1/PTGS2/ADGRA2/ITGB2/MMP9/ACTA2/STC1/FOXF1/ACVRL1/TEK/EPB41L4B/PTP4A3	27
BP	GO:00 01503	ossification	5.58E-04	SEMA7A/FAM20C/IGFBP5/MMP2/COL6A1/SH3PXD2B/ISG15/RASSF2/CSF1/CTNNBIP1/DCHS1/ADRB2/TMEM64/TGFB1/PTGS2/VCAN/CHSY1/EIF2AK3/CDH11/PKDCC/CCL3/MMP9/WNT10B/TUFT1/IFITM1/CTHRC1/STC1/ACHE/IGFBP3/PTGER4/IL6/AXIN2/CCR1/WNT7B/GLI1/TEK	36

BP	GO:00 01655	urogenital system development	5.74E-04	GDNF/SLIT2/TGFBR1/RRM2B/ALDH1A2/HAS2/CTNNBIP1/MMP17/DCHS1/AMPD2/SULF1/TGFB1/SIX1/FOXD1/SULF2/LIF/MME/RET/EPHB3/ACTA2/NOTCH3/FOXF1/WNT7B/STRA6/PDGFD/CDKN1C/EPHA7/NID1/KLF15/TEK/KIF26B/WNT6	32
BP	GO:00 30198	extracellular matrix organization	5.74E-04	MMP3/COL6A2/DPP4/TGFBR1/MMP2/GAS6/COL6A1/SH3PXD2B/HAS2/COL7A1/LRP1/PXDN/CRISPLD2/COL16A1/FURIN/SULF1/TGFB1/MMP15/ITGB2/VCAN/ADAMTS2/SULF2/POSTN/COL8A1/MMP9/ABI3BP/LOXL1/FOXF1/TNFRSF11B/NID1/MMP11/ELF3	32
BP	GO:00 35239	tube morphogenesis	5.74E-04	GDNF/SLIT2/ADAMTS12/SEMA3E/TULP3/ETV4/CSF1/CTNNBIP1/DCHS1/TCF7/TGFB1/SIX1/FOXD1/MICAL2/NFATC4/FZD2/EGF/CTHRC1/RET/EPHB3/ZEB1/HOXA1/FOXF1/STRA6/ACVRL1/GLI1/EPHA7/CSF1R/KIF26B/EDNRA/WNT6/NOTCH4/TRIM71/NTN1	34
BP	GO:00 32963	collagen metabolic process	5.74E-04	MMP3/COL6A2/MMP2/COL6A1/TRAM2/COL7A1/FURIN/TGFB1/MMP15/ADAMTS2/COL8A1/MMP9/F2R/CIITA/IL6/MMP11/COL15A1	17
BP	GO:00 43062	extracellular structure organization	5.74E-04	MMP3/COL6A2/DPP4/TGFBR1/MMP2/GAS6/COL6A1/SH3PXD2B/HAS2/COL7A1/LRP1/PXDN/CRISPLD2/COL16A1/FURIN/SULF1/TGFB1/MMP15/ITGB2/VCAN/ADAMTS2/SULF2/POSTN/COL8A1/MMP9/ABI3BP/LOXL1/FOXF1/TNFRSF11B/NID1/MMP11/ELF3	32
BP	GO:00 44236	multicellular organism metabolic process	5.94E-04	MMP3/COL6A2/MMP2/COL6A1/TRAM2/COL7A1/ADRB2/FURIN/TGFB1/MMP15/ADAMTS2/COL8A1/MMP9/F2R/CIITA/IL6/MMP11/PPARGC1A/COL15A1	19
BP	GO:00 48871	multicellular organismal homeostasis	6.65E-04	EPAS1/TRPM2/HAS2/CSF1/AMPD2/ADRB2/TMEM64/INPP5D/HSPB1/CYP4F2/HOMER1/PTGS2/CORO1A/SLC27A1/F2R/MYO5B/EDNRB/SFN/ADCY5/SGIP1/ABAT/GCNT2/IL6/IL20RB/IL1B/LDB2/MLXIPL/TNFRSF11B/CSF1R/ABCA12/SIGLEC15/RP1L1/PPARGC1A	33
BP	GO:00 32103	positive regulation of response to external stimulus	6.82E-04	CXCL1/SLIT2/GAS6/SNX4/PDCD4/SEMA5A/CDH13/CSF1/STAT5B/IL18/MAPK13/TIAM1/CXCL8/VEGFC/HSPB1/TGFB1/PTGS2/CCL3/CCL5/THBD/DDX60/PTGER4/IL6/CCR1/IL1B/PDGFD/C3/IL23A	28
BP	GO:00 02521	leukocyte differentiation	7.07E-04	IL11/FAM20C/LFNG/PIK3CD/CSF3/OSTM1/TRPM2/GAS6/LYN/RASSF2/CSF1/CTNNBIP1/STAT5B/IL18/RSAD2/TMEM64/INHBA/INPP5D/TCF7/POU2F2/TGFB1/ADA/DUSP10/MITF/LIF/CCL3/MMP9/ZFPM1/FZD8/FAS/ZEB1/PTGER4/IL6/MAFB/DOCK2/CCR1/IRF7/PLCL2/CSF1R/SIGLEC15/IL23A	41
BP	GO:00 32844	regulation of homeostatic process	7.35E-04	PIK3CD/TRPM2/PIK3CB/ISG15/LYN/NEK7/TPCN1/STAT5B/TMEM64/TIAM1/INHBA/INPP5D/IFI6/TGFB1/ADA/GPD1L/PTGS2/CORO1A/SLC27A1/ITPR1/ATP1A2/CCL3/ZFPM1/F2R/CACNA1G/JPH4/SPX/EDNRB/ADCY5/SGIP1/SYBU/ABAT/IL6/MAFB/IL1B/MLXIPL/TNFRSF11B/CSF1R/SIGLEC15/PPARGC1A	40
CC	GO:00 05578	proteinaceous extracellular matrix	3.84E-08	ADAMTSL1/MMP3/TIMP3/COL6A2/MMP2/CCBE1/ADAMTS12/LTBP2/COL6A1/WNT5B/COL7A1/TGFBR3/PXDN/MEGF9/MMP17/CRISPLD2/COL16A1/TGFB1/VCAN/ADAMTS2/POSTN/COL8A1/MMP9/WNT10B/CTHRC1/ADAMTS15/LAD1/FLRT3/ABI3BP/CD248/ACHE/LOXL1/WNT7B/APLP1/ZP1/PODNL1/TNFRSF11B/NID1/MMP11/WNT6/HMCN2/EFEMP1/COL15A1/NTN1	44
CC	GO:00 14069	postsynaptic density	7.08E-05	NPTN/SYT11/LYN/NCS1/TIAM1/DLG4/HOMER1/CAMK2A/SHANK3/SYNPO/BSN/CPEB3/CAMK2N1/ITPR1/CNIH2/GRIN2C/SHANK1/LZTS3/PDE4B/CPEB1/CRYAB/LIN7A/RGS14/GRIA1/GAP43	25
CC	GO:00 60076	excitatory synapse	2.59E-03	SYT11/DLG4/BSN/SHANK1/PDE4B/ELFN1/SYP	7
CC	GO:00 45121	membrane raft	2.59E-03	CLIP3/DPP4/TGFBR1/MYADM/CAVIN2/S100A10/CDH13/LYN/INPP5D/EFNB1/BVES/FURIN/SULF1/PTGS2/PTGIS/ATP1A2/F2R/EDNRB/FAS/RET/RTN4RL2/MAL2/TLR6/ABCB4/TEK/NOS1AP/CHRNB2	27
CC	GO:00 30425	dendrite	3.82E-03	NPTN/SYT11/CLU/KIF21B/ANXA3/NOV/LRP1/NCS1/TIAM1/DLG4/CPT1C/ADA/SHANK3/SYNPO/PLXDC1/BSN/CPEB3/SLC1A4/CAMK2N1/ATP1A2/CNIH2/SHANK1/JPH4/MME/MAPT/PDE4B/RET/EPHB3/CPEB1/RTN4RL2/ELFN1/CRYAB/RGS14/GRIA1/ACVRL1/EPHA7/PPARGC1A	37

CC	GO:00 05911	cell-cell junction	1.21E-02	CLDN11/CD2AP/KRT18/DPP4/TGFBR1/KRT8/MYADM/TMEM6 5/LYN/CGN/STXBP6/CCND1/NOV/TIAM1/BVES/SSX2IP/PKP2/ SYNPO/PLXDC1/CORO1A/ATP1A2/GJB3/NECTIN4/KIAA1462/ PANX2/FLRT3/LIN7A/CADM3/ABCB4/TEK/EPB41L4B/TRIM29	32
CC	GO:00 98589	membrane region	1.38E-02	CLIP3/DPP4/TGFBR1/MYADM/CAVIN2/S100A10/CDH13/LYN/ LRP1/INPP5D/EFNB1/BVES/FURIN/SULF1/PTGS2/PTGIS/ATP1 A2/F2R/EDNRB/FAS/SGIP1/RET/RTN4RL2/MAL2/TLR6/ABCB4 /TEK/NOS1AP/CHRNB2	29
MF	GO:00 19838	growth factor binding	5.14E-05	IGFBP4/CRIM1/IGFBP5/SHC1/NRP2/TGFBR1/LTBP2/COL6A1/ NOV/TGFBR3/PXDN/RHBDF2/FURIN/IGFBP2/IGFBP3/IL1R1/ WISP1/ACVRL1/TEK/ESM1/VASN	21
MF	GO:00 70851	growth factor receptor binding	2.39E-03	IL11/CD2AP/FGF5/NPTN/SHC1/CSF3/SNX4/LYN/VEGFC/EGF/F LRT3/TGFA/IL6/IL1R1/IL1B/PDGFD/ESM1/EFEMP1	18
MF	GO:00 04435	phosphatidylinositol phospholipase C activity	2.39E-03	BDKRB2/CCL5/PLCD1/CCR1/PLCL2/PLCH1/PLCB1/EDNRA	8
MF	GO:00 05520	insulin-like growth factor binding	2.42E-03	IGFBP4/CRIM1/IGFBP5/NOV/IGFBP2/IGFBP3/WISP1/ESM1	8
MF	GO:00 19199	transmembrane receptor protein kinase activity	3.01E-03	CRIM1/NPTN/NRP2/TGFBR1/TGFBR3/LTK/RET/EPHB3/ACVR L1/EPHA7/TEK/CSF1R/EFEMP1	13
MF	GO:00 04222	metalloendopeptidas e activity	8.97E-03	MMP3/PAPPA/MMP2/ADAMTS12/MMP17/ADAM33/CLCA2/M MP15/ADAM11/ADAMTS2/MMP9/MME/ADAMTS15/ADAM23/ MMP11	15
MF	GO:00 19955	cytokine binding	1.26E-02	NRP2/TGFBR1/TGFBR3/PXDN/TNFRSF9/FAS/IL20RB/IL1R1/C CR1/GBP1/ACVRL1/CSF1R/VASN	13
MF	GO:00 04714	transmembrane receptor protein tyrosine kinase activity	2.18E-02	CRIM1/NPTN/NRP2/LTK/RET/EPHB3/EPHA7/TEK/CSF1R/EFE MP1	10
MF	GO:00 08237	metallopeptidase activity	2.21E-02	ADAMTSL1/MMP3/PAPPA/MMP2/ADAMTS12/TRHDE/MMP17 /ADAM33/CLCA2/CPA4/MMP15/ADAM11/ADAMTS2/MMP9/M ME/ADAMTS15/ADAM23/NAALAD2/MMP11	19
MF	GO:00 04860	protein kinase inhibitor activity	3.22E-02	CDKN1A/HSPB1/CAMK2N1/SFN/LRRC15/RTN4RL2/FLRT3/TR IB2/CISH/CDKN1C/PODNL1/FAM212A	12
MF	GO:00 17154	semaphorin receptor activity	4.22E-02	NRP2/PLXNB3/PLXNA4/PLXNC1	4
MF	GO:00 97110	scaffold protein binding	4.97E-02	KRT18/KRT8/DLG4/SHANK3/SHANK1/CACNA1G/ADCY5/TRE M1	8

Go enrichment of DE genes upon NR2F2 downregulation in WI38VA13_2RA cells. Differentially expressed genes in WI38VA13_2RA cell lines (fold change >2, FDR <0.05) were mapped to GO terms using David platform, where ALL GO terms were tested. The enrichment results was summarised by allowed medium similarity=0.7 using REViGO online tool (Supek *et al.*, 2011). GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.

APPENDIX 4.7 GO enrichment of DE genes upon NR2F2 downregulation in U2OS cells

GO	ID	Description	p.adjust	GeneID	Count
BP	GO:0007059	chromosome segregation	2.47E-30	TRIP13/PTTG1/REC8/SPDL1/KIF22/CDCA8/AURKB/CDCA5/FANCD2/RACGAP1/TACC3/CENPM/BRCA1/EME1/ERCC6L/SKA1/GSG2/FAM83D/CDCA2/CENPW/KIF2C/PSRC1/KIF4A/ZWINT/KIFC1/SKA3/HJURP/NCAPH/CENPK/NDC80/CCNB1/FEN1/PRC1/CENPF/ESPL1/CENPE/NCAPG/BIRC5/SGO2/UBE2C/CDT1/KIF18A/KIF18B/BUB1B/SPC25/CDC20/PLK1/KIF11/MKI67/NUSAP1/NUF2/NEK2/BUB1/KIF14/CENPU/ESCO2/KNL1/CENPA/TOP2A/SPAG5/MAD2L1/SPC24/OIP5/BRIP1/TTK/SGO1/DSCC1/DLGAP5/DMC1	69
BP	GO:0007067	mitotic nuclear division	2.99E-27	MISP/PTTG1/SPDL1/CDC25A/KIF22/CDCA8/AURKB/CDCA5/RACGAP1/TACC3/CIT/PKMYT1/ERCC6L/SKA1/CCNF/TPX2/GSG2/BORA/AURKA/FAM83D/CDCA2/CENPW/KIF2C/PSRC1/CCNB2/PIMREG/KIF4A/KIF15/ZWINT/KIFC1/CDCA3/SKA3/NCAPH/HAUS8/NDC80/CCNB1/PRC1/CENPF/ESPL1/CENPE/CDC25C/NCAPG/BIRC5/MYBL2/KIF20B/UBE2C/CDT1/CCNA2/KIF18A/KIF18B/BUB1B/SPC25/CDC20/PLK1/KIF11/EREG/MKI67/NUSAP1/NUF2/NEK2/BUB1/KIF14/ASPM/CDK1/KNL1/CENPA/SPAG5/MAD2L1/PDGFRB/SPC24/OIP5/TTK/SGO1/DSCC1/PBK/DLGAP5	76
BP	GO:0006260	DNA replication	3.48E-10	RRM2B/MCM7/CDC25A/ACHE/GINS3/RFC4/BLM/RFC3/PIF1/BRCA1/CDC45/EME1/TICRR/RRM1/MCM2/MCM10/POLQ/FEN1/MCM5/CDC25C/PCLAF/CDT1/ORC1/CHAF1A/CHAF1B/EXO1/MCM4/EREG/RAD51/DTL/BMP6/ESCO2/CDK1/MCM6/BRIP1/DSCC1/RRM2/GINS2/E2F8	39
BP	GO:0000226	microtubule cytoskeleton organization	6.50E-09	MAP1A/CLIP3/SPDL1/TRIM54/AURKB/RACGAP1/TACC3/BRCA1/DISC1/SKA1/CCNF/TPX2/BORA/AURKA/XRCC2/KIF2C/PSRC1/KIF4A/KIFC1/SKA3/MELK/HAUS8/NDC80/CCNB1/PRC1/AUNIP/ESPL1/BIRC5/MYBL2/PCLAF/KIF18A/KIF18B/SPC25/CDC20/PLK1/KIF11/PLK4/NUSAP1/NEK2/ASPM/CDK1/CENPA/SPAG5/TTK/SGO1/KIF20A/DEUP1	47
BP	GO:0051321	meiotic cell cycle	4.94E-08	TRIP13/PTTG1/REC8/FANCD2/FANCA/EME1/AURKA/XRCC2/NCAPH/MND1/ESPL1/H2AFX/SGO2/KIF18A/BUB1B/CDC20/PLK1/EXO1/EREG/MKI67/RAD51/NEK2/BUB1/ASPM/CDK1/TOP2A/MOV10L1/BRIP1/TTK/SGO1/DMC1	31
BP	GO:0000075	cell cycle checkpoint	5.59E-08	SPDL1/AURKB/BLM/BRCA1/CDC45/EME1/GSG2/TICRR/AURKA/SOX11/FANCI/ZWINT/NDC80/CCNB1/CENPF/H2AFX/CDC25C/CDT1/BUB1B/PLK1/DTL/BUB1/E2F1/CDK1/KNL1/TOP2A/MAD2L1/BRIP1/TTK/TP73	30
BP	GO:0006261	DNA-dependent DNA replication	8.79E-08	RRM2B/MCM7/GINS3/RFC4/BLM/RFC3/CDC45/EME1/TICRR/MCM2/MCM10/POLQ/FEN1/MCM5/CDT1/ORC1/MCM4/RAD51/BMP6/MCM6/DSCC1/GINS2/E2F8	23
BP	GO:0034501	protein localization to kinetochore	1.10E-07	SPDL1/AURKB/GSG2/NDC80/CDT1/BUB1B/CDK1/KNL1/TTK	9
BP	GO:0050000	chromosome localization	2.51E-07	SPDL1/KIF22/CDCA8/CDCA5/FAM83D/KIF2C/PSRC1/KIFC1/NDC80/CCNB1/CENPF/CENPE/CDT1/KIF18A/KIF14/DLGAP5	16
BP	GO:0006270	DNA replication initiation	6.44E-06	MCM7/CDC45/TICRR/MCM2/MCM10/MCM5/CDT1/ORC1/MCM4/MCM6/GINS2	11
BP	GO:0034502	protein localization to chromosome	6.64E-05	SPDL1/AURKB/CDCA5/GSG2/NDC80/CDT1/BUB1B/PLK1/ESCO2/CDK1/KNL1/CENPA/TTK	13

BP	GO:0033044	regulation of chromosome organization	9.89E-05	MYC/OGT/AURKB/PAX5/CDCA5/TACC3/PIF1/BRCA1/CDC45/NDC80/CCNB1/FEN1/CENPF/ESPL1/CENPE/UBE2C/CDT1/BUB1B/PLK1/MKI67/NEK2/BUB1/PRDM12/MAD2L1/TTK/DLGAP5/ATAD2/NOS1	28
BP	GO:0006323	DNA packaging	1.55E-04	CDCA5/CENPM/HMGB2/MCM2/CENPW/HJURP/NCAPH/CENPK/CCNB1/H2AFX/NCAPG/ASF1B/CHAF1A/CHAF1B/NUSAP1/CENPU/PADI4/CDK1/KNL1/CENPA/TOP2A/OIP5	22
BP	GO:0071897	DNA biosynthetic process	1.96E-04	MYC/AURKB/RFC4/BLM/RFC3/PIF1/BRCA1/XRCC2/RAD51AP1/CTGF/POLQ/CENPF/PCLAF/TK1/EXO1/RAD51/DTL/NEK2/PDGFRB/BRIP1/DSCC1/KIAA2022	22
BP	GO:0000731	DNA synthesis involved in DNA repair	5.40E-04	RFC4/BLM/RFC3/BRCA1/XRCC2/RAD51AP1/PCLAF/EXO1/RAD51/DTL/BRIP1/KIAA2022	12
BP	GO:0065004	protein-DNA complex assembly	8.39E-04	WNT10B/CENPM/CDC45/HMGB2/MCM2/XRCC2/CENPW/HJURP/CENPK/CENPF/H2AFX/CENPE/CDT1/ASF1B/CHAF1A/CHAF1B/RAD51/CENPU/PADI4/KNL1/CENPA/OIP5/DMC1	23
BP	GO:0000732	strand displacement	1.33E-03	BLM/BRCA1/XRCC2/RAD51AP1/EXO1/RAD51/BRIP1	7
BP	GO:0051383	kinetochore organization	1.33E-03	CENPW/NDC80/CENPF/CENPE/CDT1/CENPA	6
BP	GO:0035825	reciprocal DNA recombination	1.45E-03	TRIP13/REC8/EME1/XRCC2/MND1/RAD51/TOP2A/BRIP1/DMC1	9
CC	GO:0000793	condensed chromosome	1.90E-21	REC8/SPDL1/AURKB/CDCA5/FANCD2/BLM/CENPM/BRCA1/ERCC6L/SKA1/HMGB2/AURKA/SUV39H1/CENPW/KIF2C/ZWINT/SKA3/HJURP/NCAPH/CENPK/NDC80/CCNB1/CENPF/H2AFX/CENPE/NCAPG/BIRC5/SGO2/CDT1/BUB1B/SPC25/PLK1/MKI67/RAD51/NUF2/NEK2/BUB1/CENPU/KNL1/CENPA/TOP2A/SPAG5/MAD2L1/SPC24/SGO1/DMC1	46
CC	GO:0098687	chromosomal region	2.71E-20	REC8/SPDL1/MCM7/KIF22/CDCA8/AURKB/CDCA5/BLM/CENPM/PIF1/ERCC6L/SKA1/AURKA/MCM2/SUV39H1/CENPW/KIF2C/ZWINT/SKA3/HJURP/CENPK/NDC80/CCNB1/FEN1/MCM5/CENPF/H2AFX/CENPE/NCAPG/BIRC5/SGO2/CDT1/KIF18A/ORC1/BUB1B/SPC25/PLK1/MCM4/MKI67/RAD51/NUF2/NEK2/BUB1/CENPU/ESCO2/CDK1/KNL1/CENPA/SPAG5/MAD2L1/SPC24/MCM6/OIP5/TTK/SGO1/DSCC1/DMC1	57
CC	GO:0005819	spindle	5.29E-16	STX1B/SPDL1/KIF22/CDCA8/AURKB/RACGAP1/CEP128/TACC3/CEP19/SKA1/TPX2/GSG2/AURKA/FAM83D/PSRC1/KIF4A/KIF15/KIFC1/SKA3/CKAP2L/HAUS8/CCNB1/PRC1/AUNIP/CENPF/ESPL1/CENPE/BIRC5/KIF20B/KIF18A/KIF18B/BUB1B/CDC20/PLK1/KIF11/SHCBP1/NUSAP1/NEK2/KIF14/ASPM/CDK1/SPAG5/MAD2L1/TTK/SGO1/DLGAP5/KIF20A	47
CC	GO:0005871	kinesin complex	1.46E-08	KIF22/DISC1/KIF5C/KIF2C/KIF4A/KIF15/KIFC1/CENPE/KIF20B/KIF18A/KIF18B/KIF11/KIF14/KIF26A/KIF20A	15
CC	GO:0030496	midbody	5.37E-06	CDCA8/AURKB/RACGAP1/AURKA/CEP55/PSRC1/KIF4A/PRC1/CENPF/CENPE/BIRC5/KIF20B/PLK1/SHCBP1/NEK2/KIF14/ASPM/CDK1/SPAG5/KIF20A	20
CC	GO:0005578	proteinaceous extracellular matrix	4.20E-05	FN1/PXDN/MFAP5/LUM/WNT10A/ADAMTS12/SPON2/EPYC/COL24A1/ACHE/WNT10B/TIMP3/COL17A1/LINGO1/WNT6/LRRN1/FGF1/WNT5B/ADAMTS5/CTGF/APLP1/ZP3/HAPLN2/ADAMTSL1/COL5A3/LAMC2/LAMA4/IMPG2/TNC/WNT7B/MMP3/COL3A1	32
CC	GO:0042555	MCM complex	4.51E-04	MCM7/MCM2/MCM5/MCM4/MCM6	5
CC	GO:0030018	Z disc	1.45E-03	ANK2/TRIM54/ANK1/FBXL22/SYNPO2L/MYBPC2/SYNC/IGSF22/CSRP3/PALLD/NOS1/ACTN2/MYOZ2/XIRP2	14

CC	GO:0033267	axon part	7.80E-03	SYT11/L1CAM/SYP/SYT7/ANK1/SYT1/PVALB/KCNQ2/TPX2/AURKA/KIF4A/NMU/CPLX1/KCNA1/SCN2A/CALB1	16
CC	GO:0000800	lateral element	1.13E-02	REC8/BLM/BRCA1/RAD51	4
CC	GO:0010369	chromocenter	1.45E-02	CDCA8/AURKB/ESCO2/OIP5	4
CC	GO:0043073	germ cell nucleus	3.48E-02	TRIP13/REC8/AURKA/H2AFX	4
CC	GO:0005581	collagen trimer	4.18E-02	LUM/COL24A1/COL17A1/COL5A3/COLEC12/C1QL2/COLEC10/MARCO/COL3A1	9
MF	GO:0008017	microtubule binding	8.52E-06	MAP1A/CLIP3/TRIM54/KIF22/SYBU/RACGAP1/KIF5C/SKA1/FAM83D/KIF2C/PSRC1/KIF4A/KIF15/KIFC1/PRC1/CENPE/BIRC5/KIF20B/KIF18A/KIF18B/PLK1/KIF11/NUSAP1/KIF14/KIF26A/KIF20A/JAKMIP2	27
MF	GO:0008094	DNA-dependent ATPase activity	7.22E-05	MCM7/RFC4/BLM/RFC3/PIF1/XRCC2/POLQ/MCM4/RAD51/TOP2A/MCM6/BRIP1/DSCC1/DMC1	14
MF	GO:0003777	microtubule motor activity	7.22E-05	KIF22/KIF5C/KIF2C/KIF4A/KIF15/KIFC1/CENPE/KIF20B/KIF18A/KIF18B/KIF11/KIF14/KIF26A/KIF20A	14
MF	GO:0016887	ATPase activity	9.66E-05	MCM7/KIF22/ATP8A2/RFC4/BLM/RFC3/PIF1/KIF5C/XRCC2/KIF2C/KIF4A/KIF15/KIFC1/POLQ/CENPE/KIF20B/KIF18A/KIF18B/KIF11/MCM4/RAD51/ABCA12/KIF14/KIF26A/ABCB5/TOP2A/MOV10L1/ABCG2/MCM6/BRIP1/DSCC1/ABCD2/KIF20A/ATAD2/DMC1/ATP10B	36
MF	GO:0043142	single-stranded DNA-dependent ATPase activity	1.66E-04	RFC4/RFC3/PIF1/POLQ/RAD51/DSCC1	6
MF	GO:0003678	DNA helicase activity	7.73E-04	MCM7/BLM/PIF1/CDC45/MCM2/MCM5/MCM4/MCM6/BRIP1/GINS2	10
MF	GO:0000217	DNA secondary structure binding	8.42E-04	BLM/HMGB2/XRCC2/RAD51/NR0B1/DMC1/NEIL3	7
MF	GO:0008574	ATP-dependent microtubule motor activity, plus-end-directed	8.49E-04	KIF5C/KIF4A/KIF20B/KIF18A/KIF11/KIF14	6
MF	GO:0038024	cargo receptor activity	1.20E-03	TFR2/SSC4D/LRP1/MEGF10/CD163/ENPP2/COLEC12/MARCO/FOLR1/LOXL4/TMPRSS2	11
MF	GO:0072349	modified amino acid transmembrane transporter activity	1.52E-03	SLC1A4/SLCO4A1/SLC7A11/SLCO1C1/SLCO1B1/SLC16A12	6

Go enrichment of DE genes upon NR2F2 downregulation in U2OS cells. Differently expressed genes in U2OS cell lines (fold change >4, FDR <0.05) was mapped to GO terms using David platform, where ALL GO terms were tested. The enrichment results was summarised by allowed medium similarity=0.7 using REViGO online tool (Supek *et al.*, 2011). GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.

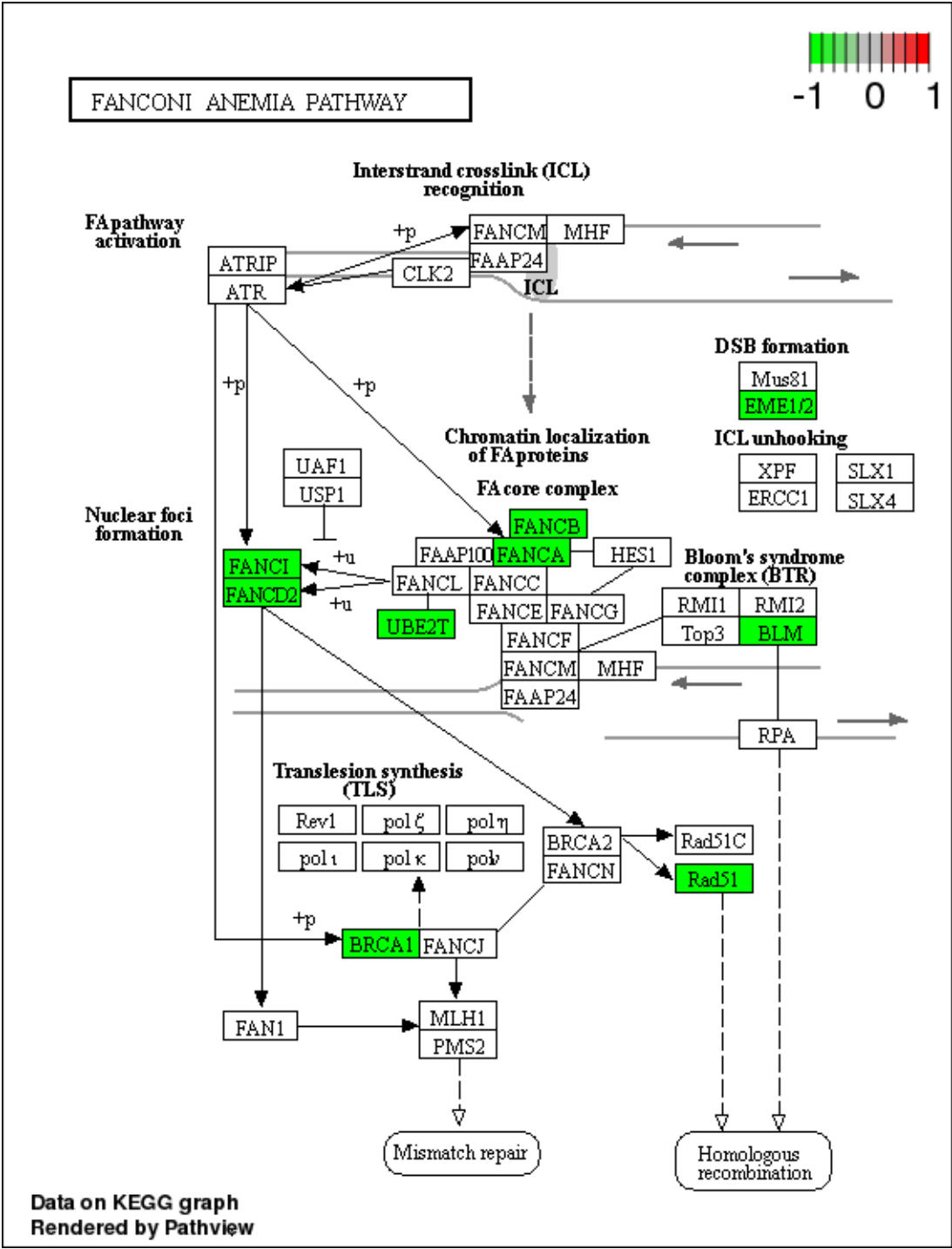**APPENDIX 4.8 Visualising the DE genes upon NR2F2 depletion in selected KEGG pathways in U2OS cells**

Visualising the DE genes upon NR2F2 depletion in FA pathway in U2OS cells.

Visualising the DE genes upon NR2F2 depletion in DNA replication in U2OS cells.

Visualising the DE genes upon NR2F2 depletion in HR pathway in U2OS cells.
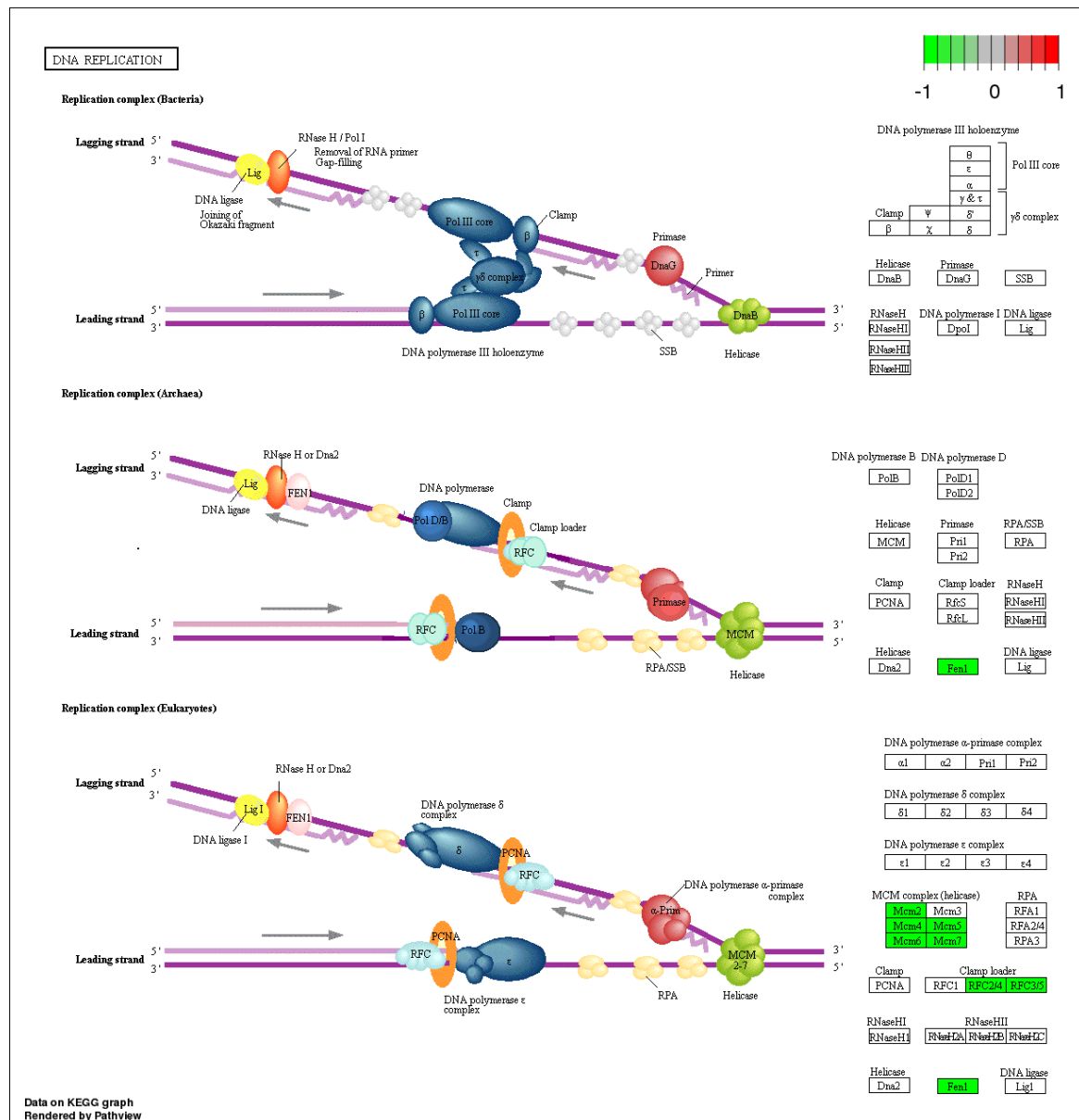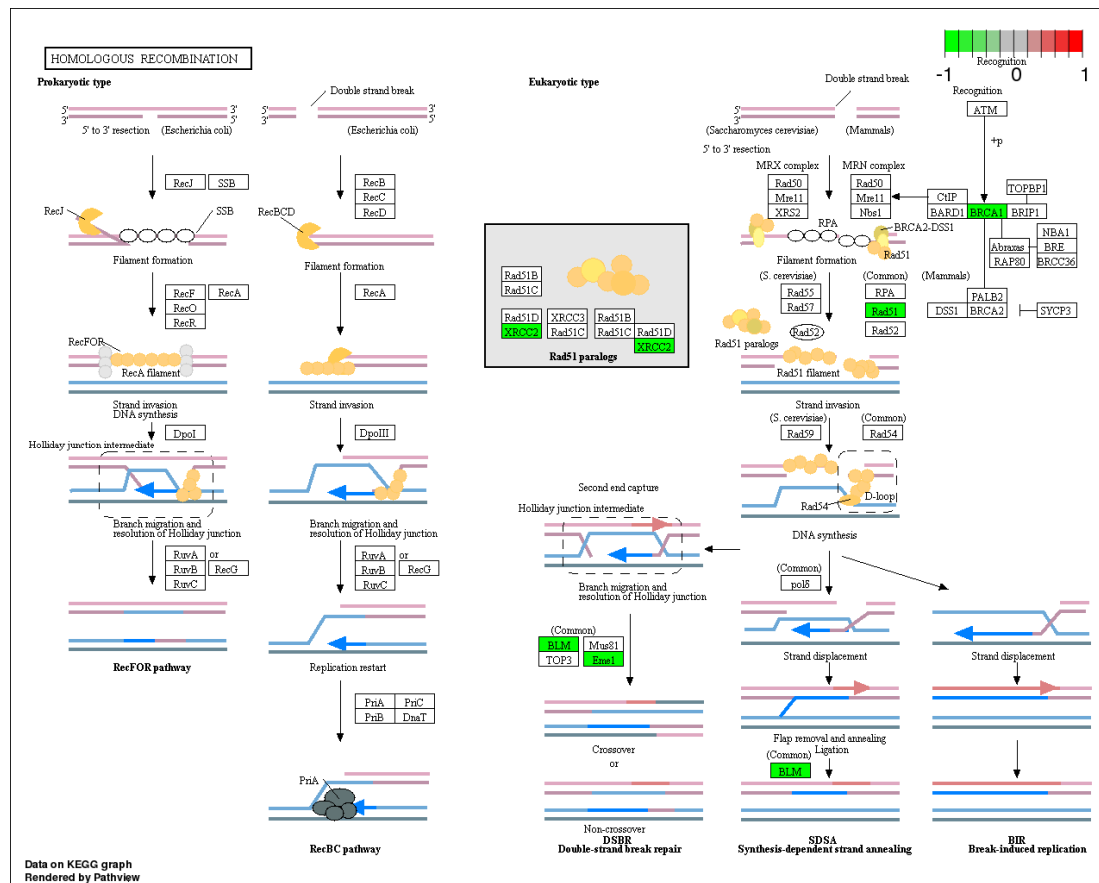
Appendixes - chapter 5

APPENDIX 5.1 PCR amplification of across the CRISPR targets in TCL1 cells

A)
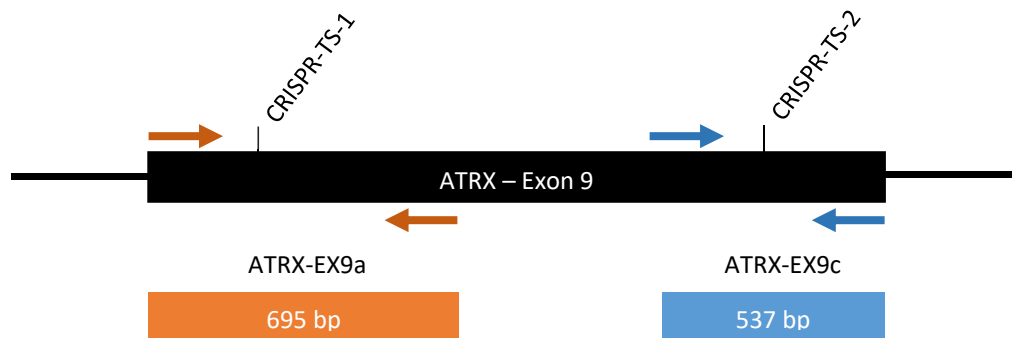
B)
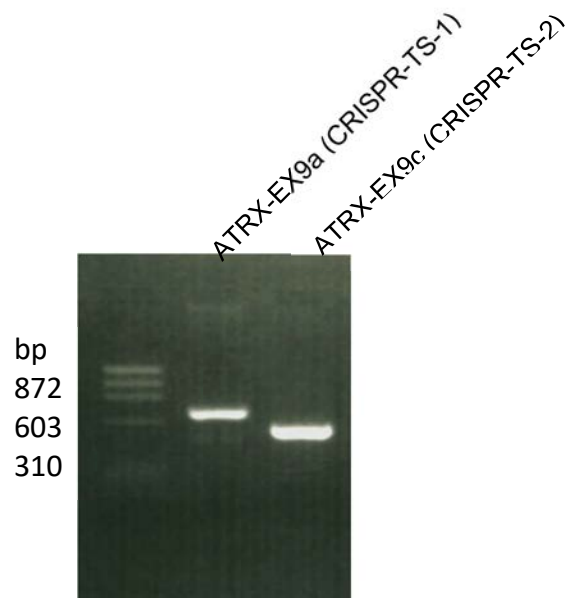
PCR amplification of across the CRISPR targets in TCL1 cells. (A) diagram showing the location of the ATRX-EX9a amplicon (primers ATRX EX9a.F and ATRX EX9a.R that generate an amplicon of 695 bp) which covers the CRISPR-TS-1 target and ATRX-Ex9c amplicon (primer ATRX EX9c.F and ATRX EX9c2.R that generate an

amplicon od 53 bp) which covers the CRISPR-TS-2 target (B) Gel electrophoresis of the ATRX EX9a and ATRX EX9a amplicons from TCL1 cells.

APPENDIX 5.2 Validation of the CRISPR target sites in ATRX exon9 loci in TCL1 cells using Sanger sequencing

A)
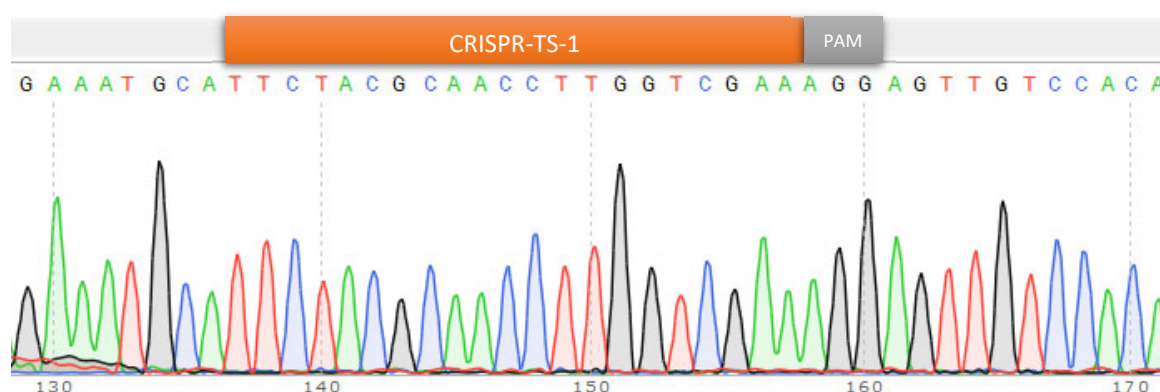
B)
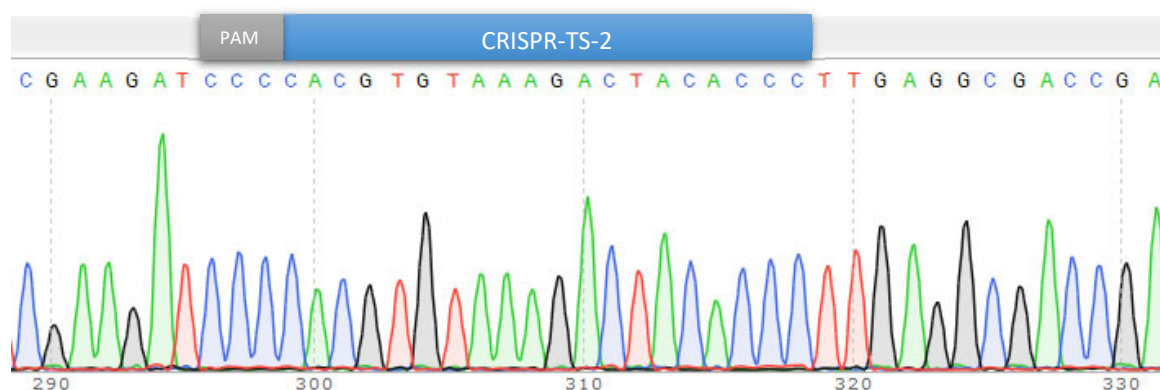
Validation of the CRISPR target sites in *ATRX* exon9 loci in TCL1 cells using Sanger sequencing. (A) Sanger sequencing chromatogram of ATRX-EX9a using ATRX EX9a.R primer CRISPR-TS-1 sequence is marked by the orange bar. (B) Sanger

sequencing chromatogram of ATRX-EX9c using ATRX EX9c.F primer. The reverse complementary sequence of CRISPR-TS-2 is marked by the blue bar.

APPENDIX 5.3 Predicted restriction digest map of the pLEIC-CRISPR vectors after cloning

Double digest

Size (bp)		site1		site2	Mass %
5973	NotI	6433	EcoRV	3222	65
3211	EcoRV	3222	NotI	6433	35

Triple digest

Size (bp)		site1		site2	Mass %
4493	PvuI	7913	EcoRV	3222	49
3211	EcoRV	3222	NotI	6433	35
1480	NotI	6433	PvuI	7913	16

The full length of CRISPR vectors after cloning are the same for pLEIC-ATRX-CRISPR-TS-1 and pLEIC-ATRX-CRISPR-TS-2, which is 9183 bp. The predicted fragment size (bp) and mass (%) were calculated using the ApE-A plasmid Editor v2.0 software.

APPENDIX 5.4 Sequencing alignment analysis of ATRX editing at CRISPR-TS-1 and CRISPR-TS-2
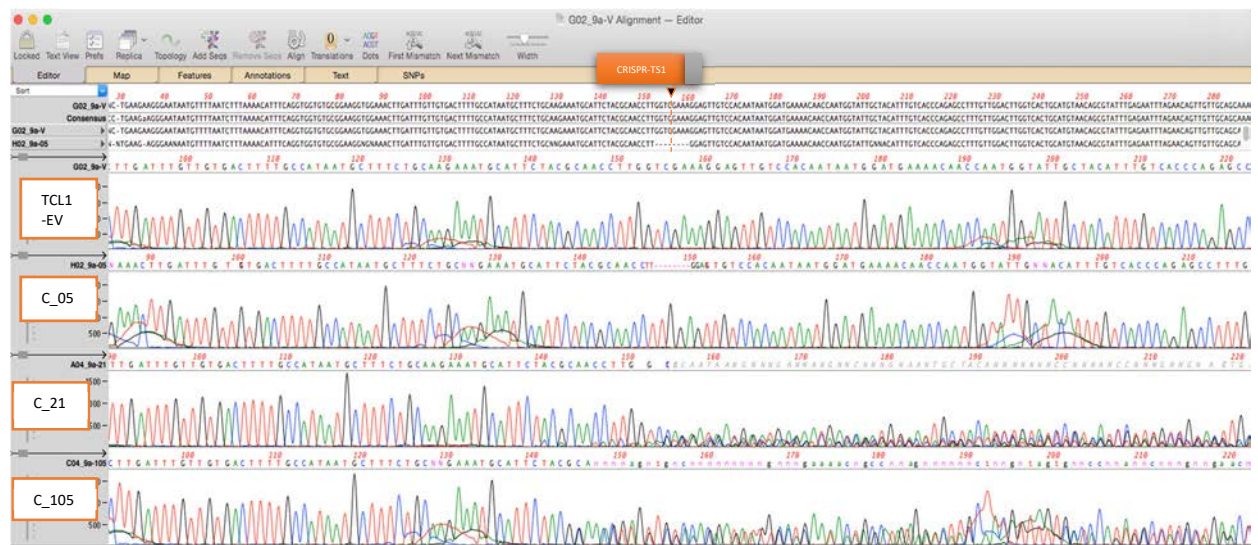
Sequencing alignment analysis of ATRX editing at CRISPR-TS-1. TCL1-EV has used as a reference sequence. The CRISPR-TS-1 cleavage site is identified in orange.
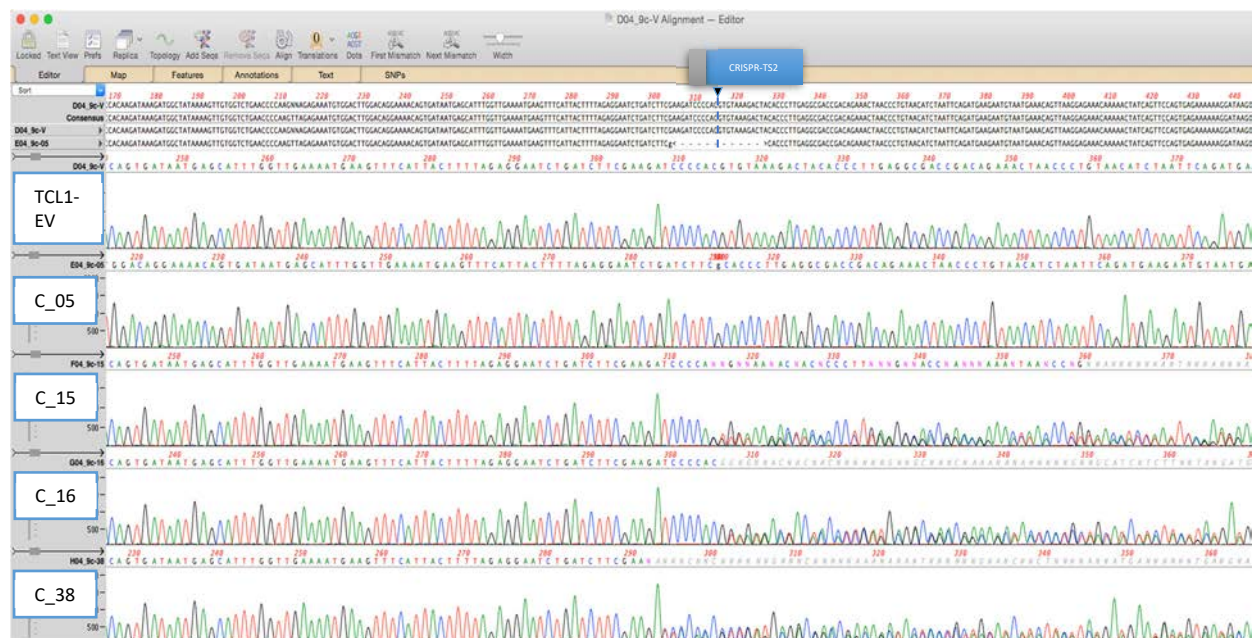
Sequencing alignment analysis of ATRX editing at CRISPR-TS-2. TCL1-EV was used as a reference. CRISPR-TS-2 cleavage site is identified in blue.

APPENDIX 5.5 Sequencing alignment analysis of ATRX editing at CRISPR-TS-1 and CRISPR-TS-2

ATRX [Homo sapiens] 2492 aa

GenBank: BAC81110.1

Domains   Zn binding sites   peptide binding sites   Exon 9

ATP binding sites   putative Mg++ binding site   nucleotide binding regions

```
ORIGIN
        1 mtaepmsesk lntlvqklhd flahsseese etsspprlam nqntdkisgs gsnsdmmens
       61 keegtsssek skssgssrsk rkpsivtkyv esddekpldd etvnedasne nsenditmqs
      121 lpkgtvivqp epvlnedkdd fkgpefrsrs kmktenlkkr gedglhgivs ctacgqqvnh
      181 fqkdsiyrhp slqvlicknc fkyymsddis rdsdgmdeqc rwcaeggnli ccdfchnafc
      241 kkcilrnlgr kelstimden nqwycyichp eplldlvtac nsvfenleql lqqnkkkikv
      301 dseksnkvye htsrfspkkt ssncngeekk lddscsgsvt ysysalivpk emikkakkli
      361 ettanmnssy vkflkqatdn seissatklr qlkafksvla dikkahlale edlnsefram
      421 davnkekntk ehkvidakfe tkarkgekpc alekkdisks eaklsrkqvd sehmhqnvpt
      481 eeqrtnkstg gehkksdrke epqyepants edldmdivsv pssvpedife nletamevqs
      541 svdhqgdgss gteqevesss vklnisskdn rggiksktta kvtkelyvkl tpvslsnspi
      601 kgadcqevpq dkdgykscgl npklekcglg qensdnehlv enevslllee sdlrrsprvk
      661 ttplrrptet npvtsnsdee cnetvkekqk lsvpvrkkdk rnssdsaidn pkpnklpksk
      721 qsetvdqnsd sdemlailke vsrmshssss dtdineihtn hktlydlktq agkddkgkrk
      781 rksstsgsdf dtkkgksaks siiskkkrqt qsessnydse lekeiksmsk igaarttkkr
      841 ipntkdfdss edekhskkgm dnqghknlkt sqegssddae rkqeretfss aegtvdkdtt
      901 imelrdrlpk kqqasastdg vdklsgkees ftslevrkva etkekskhlk tktckkvqdg
      961 lsdiaekflk kdqsdetsed dkkqskkgte ekkkpsdfkk kvikmeqqye sssdgteklp
     1021 ereeichfpk gikqikngtt dgekkskkir dktskkkdel sdyaekstgk gdscdssedk
     1081 kskngaygre kkrckllgks srkrqdcsss dtekysmked gcnssdkrlk rielrerrnl
     1141 sskrntkeiq sgssssdaee ssednkkkkq rtsskkkavi vkekkrnslr tstkrkqadi
     1201 tsssssdied ddqnsigegs sdeqkikpvt enlvlsshtg fcqssgdeal sksvpvtvdd
     1261 ddddndpenr iakkmlleei kanlssdedg ssddepeegk krtgkqneen pgdeeaknqv
     1321 nsesdsdsee skkpryrhrl lrhkltvsdg esgeekktkp kehkevkgrn rrkvssedse
     1381 dsdfqesgvs eevsesedeq rprtrsakka eleenqrsyk qkkkrrrikv qedsssenks
     1441 nseeeeeeke eeeeeeeeee eeeedendds kspgkgrkki rkilkddklr tetqnalkee
     1501 eerrkriaer erereklrev ieiedasptk cpittklvld edeetkeplv qvhrnmvikl
     1561 kphqvdgvqf mwdcccesvk ktkkkspgsgc ilahcmglgk tlqvvsflht vllcdkldfs
     1621 talvvcplnt alnwmnefek wqeglkddek levselatvk rpqersymlq rwqedggvmi
     1681 igyemyrnla qgrnvksrkl keifnkalvd pgpdfvvcde ghilkneasa vskamnsirs
     1741 rrriiltgtp lqnnlieyhc mvnfikenll gsikefrnrf inpiqngqca dstmvdvrvm
     1801 kkrahilyem lagcvqrkdy taltkflppk heyvlavrmt siqcklyqyy ldhltgvgnn
     1861 seggrgkaga klfqdfqmls riwthpwclq ldyiskenkg yfdedsmdef iasdsdetsm
     1921 slssddytkk kkkgkkgkkd ssssgsgsdn dvevikvwns rsrgggegnv detgnnpsvs
     1981 lkleeskats ssnpsspapd wykdfvtdad aevlehsgkm vllfeilrma eeigdkvlvf
     2041 sqslisldli edflelasre ktedkdkpli ykgegkwlrn idyyrldgst taqsrkkwae
     2101 efndetnvrg rlfiistkag slginlvaan rviifdaswn psydiqsifr vyrfgqtkpv
     2161 yvyrflaqgt medkiydrqv tkqslsfrvv dqqqverhft mneltelytf epdllddpns
     2221 ekkkkrdtpm lpkdtilael lqihkehivg yhehdslldh keeeelteee rkaawaeyea
     2281 ekkgltmrfn iptgtnlppv sfnsqtpyip fnlgalsams nqqledlinq grekvveatn
     2341 svtavriqpl ediisavwke nmnlseaqvq alalsrqasq eldvkrreai yndvltkqqm
     2401 liscvqrilm nrrlqqqynq qqqqqmtyqq atlghlmmpk ppnlimnpsn yqqidmrgmy
     2461 qpvaggmqpp plqrapppmr sknpgpsqgk sm
```

ADDz domain, PHD-like zinc finger domain

encodes by Exon 9 (222-1245)

SNF2 N domain

Helicase superfamily c-terminal domain

ATRX Ab-immunogen sequence Sigma, H-300

Appendixes - chapter 6

APPENDIX 6.1 Metrics of RNA-seq data from pre-crisis and post-crisis
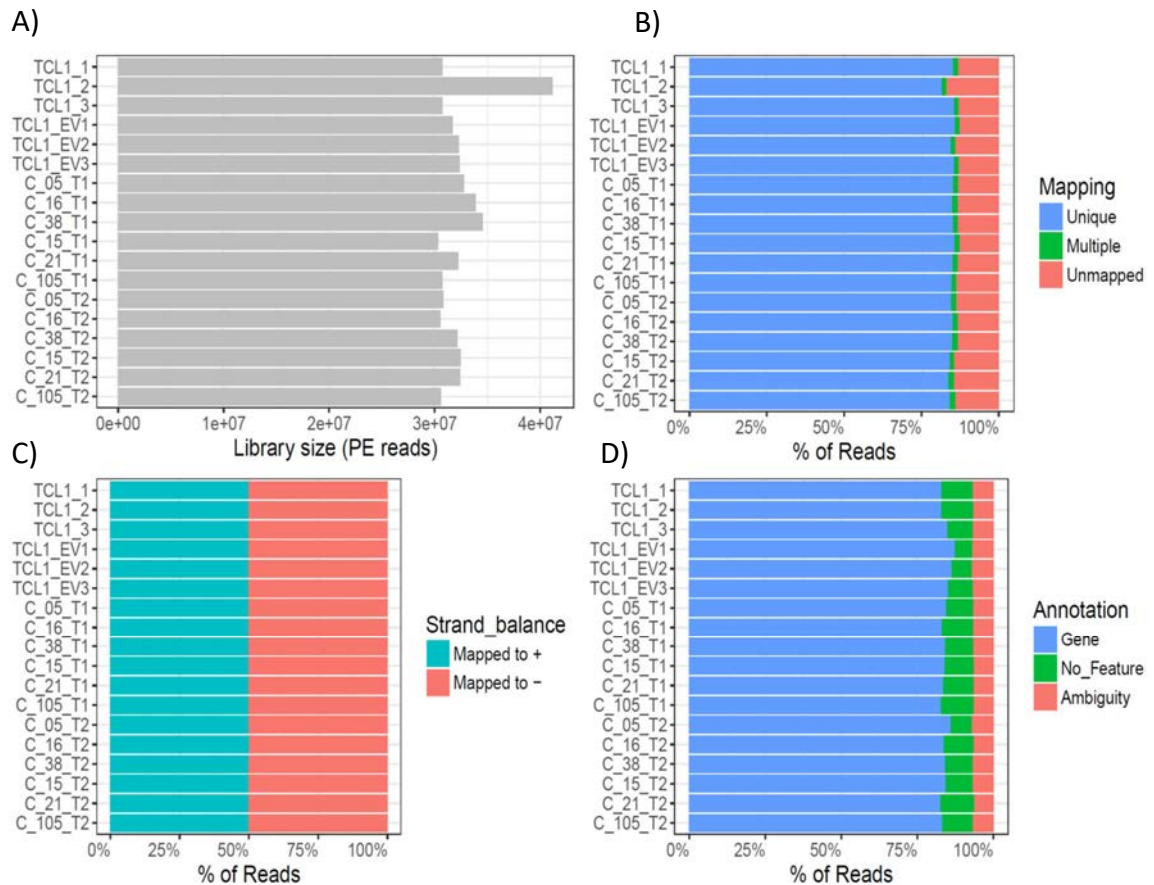
Metrics of RNA-seq data from pre-crisis and post-crisis. (A) the sequencing library size represented by total number of PE clean reads, (B) The mapping summaries for PE clean reads. Unique- represent the uniquely mapped PE sequences. Multiple- is the fraction of PE reads that has mapped to more than one site, Unmapped- represent the fraction of number of reads that has not mapped at all, (C) The strand balance of mapped reads, (D) Counting of uniquely mapped PE reads. It shows that about 83-87% of all reads were uniquely mapped to genes (known features).

APPENDIX 6.2 Gene ontology enrichment of DE genes between HGI_T1 vs TCL1
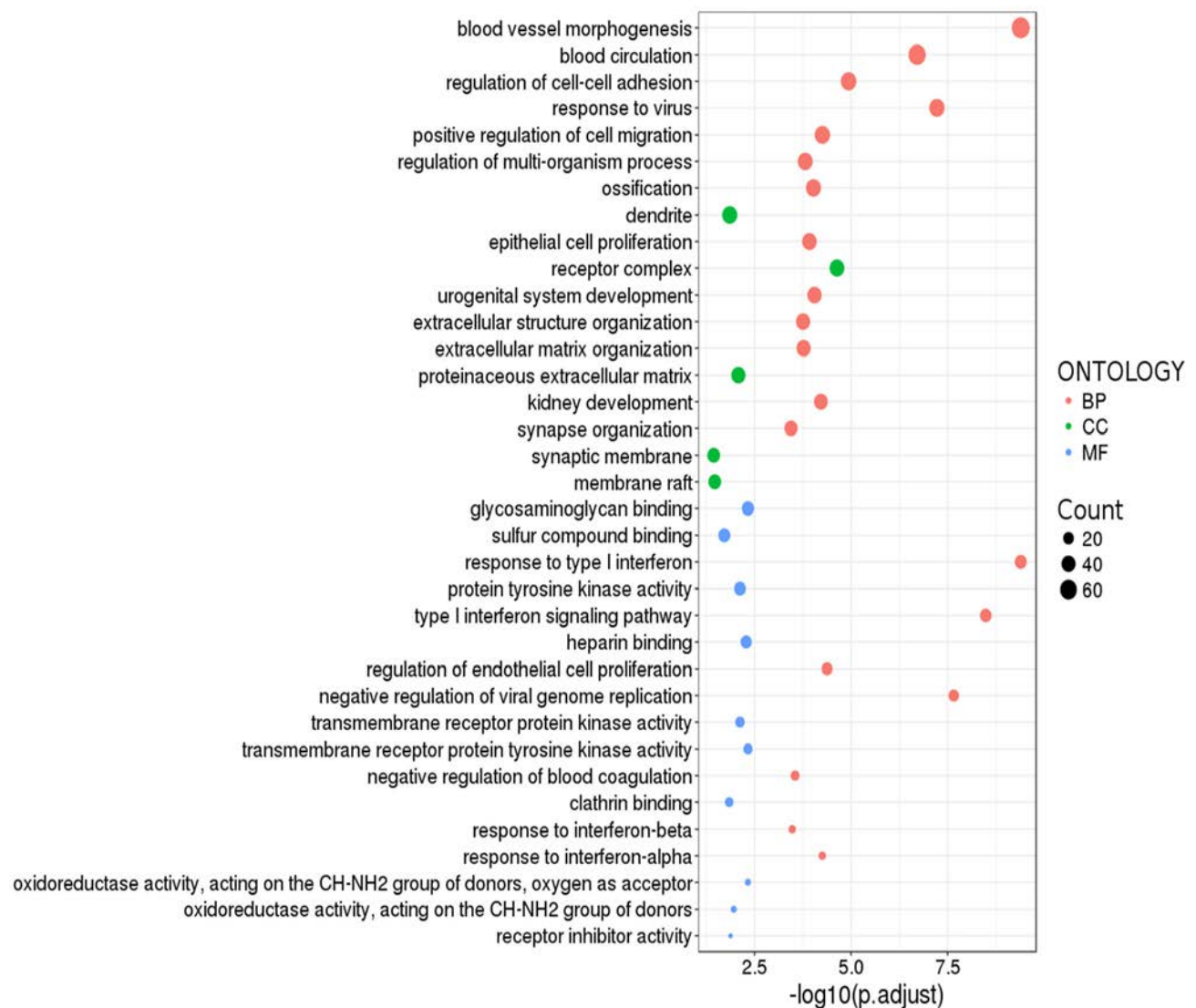
Go enrichment of DE genes between HGI_T1 vs TCL1. It illustrates the GO enrichment of DE genes (fold change >2, FDR <0.05), where ALL GO terms were tested. GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.

APPENDIX 6.3 Gene ontology enrichment of DE genes between LGI_T1 vs TCL1
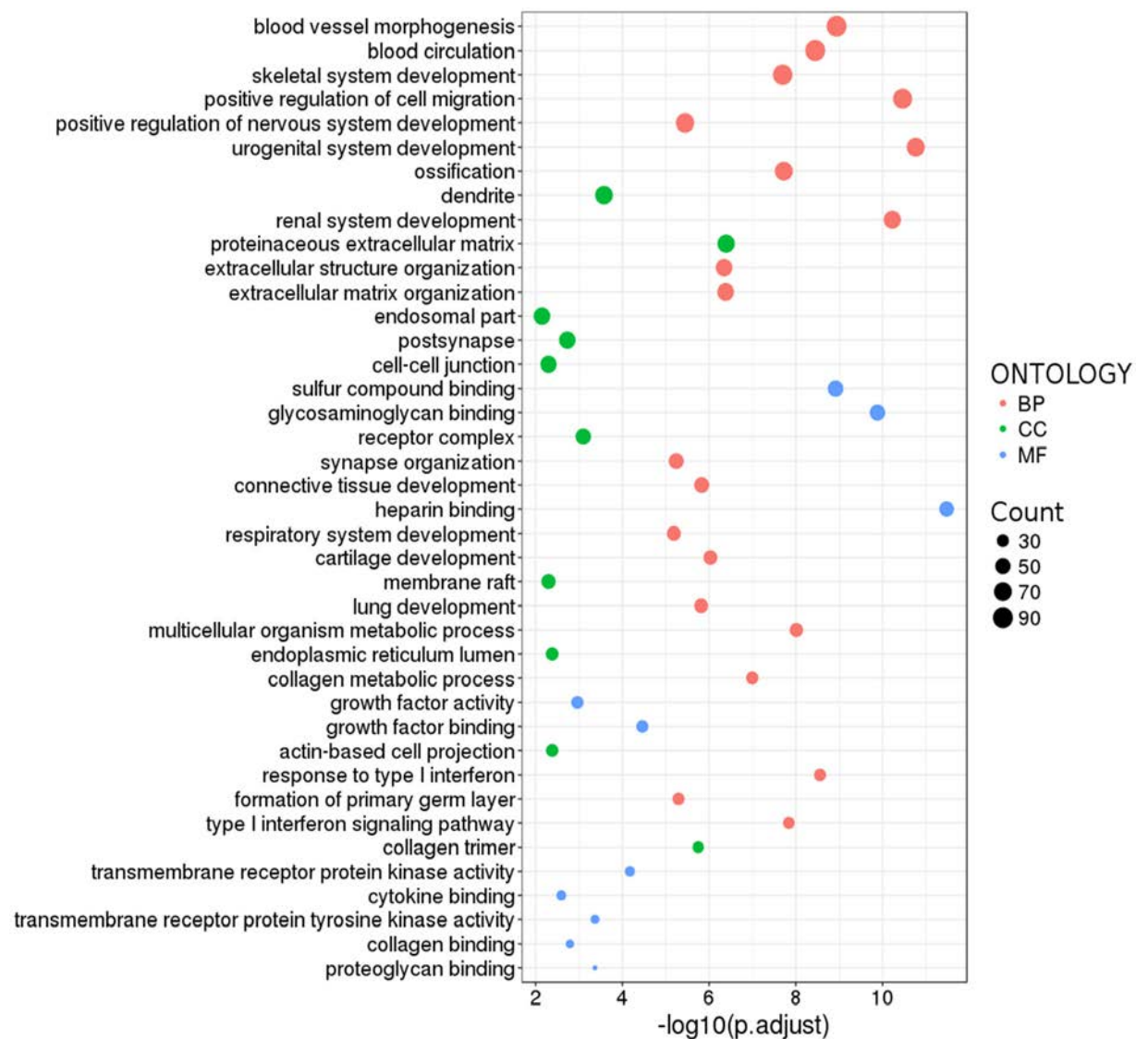
Gene ontology enrichment of DE genes between LGI_T1 vs TCL1. It illustrates the GO enrichment of DE genes (fold change >2, FDR <0.05), where ALL GO terms were tested. GO terms: BP = Biological Process, CC= Cellular components, MF=Molecular function.

APPENDIX 6.4 Comparative GO enrichment analysis of DE genes in subgroups at T1 and T2

Cluster	GO ID	Description	Gene Ratio	Bg Ratio	pvalue	p.adjust	qvalue	Gene ID	Count
downregulated.HGI_T1	GO:0016641	oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	5/574	15/16982	9.82E-05	0.054336	0.051282	IL4I1/LOXL4/LOX/LOXL3/AOC3	5
downregulated.HGI_T1	GO:0001664	G-protein coupled receptor binding	21/574	254/16982	0.000154	0.054336	0.051282	CCL5/CXCL11/CCL2/PALM/GAL/CCL7/CCL28/ADA2/RTP4/PYY2/CTHRC1/ARRB1/CCL26/APLN/ASIP/AGTR1/PDE4D/SDCBP/PSAP/CXCL16/WNT2B	21
downregulated.HGI_T1	GO:0004714	transmembrane receptor protein tyrosine kinase activity	9/574	64/16982	0.000284	0.055609	0.052483	FLT1/TIE1/FLT4/RET/EFNB3/MST1R/TEK/ALK/NRP2	9
downregulated.HGI_T1	GO:0019199	transmembrane receptor protein kinase activity	10/574	80/16982	0.000353	0.055609	0.052483	FLT1/TIE1/FLT4/RET/EFNB3/ACVRL1/MST1R/TEK/ALK/NRP2	10
downregulated.HGI_T1	GO:0016638	oxidoreductase activity, acting on the CH-NH2 group of donors	5/574	20/16982	0.000441	0.055609	0.052483	IL4I1/LOXL4/LOX/LOXL3/AOC3	5
downregulated.HGI_T1	GO:0004859	phospholipase inhibitor activity	4/574	12/16982	0.000515	0.055609	0.052483	SNCB/APOC1/ANXA4/PINLYP	4
downregulated.HGI_T1	GO:0005507	copper ion binding	8/574	56/16982	0.000551	0.055609	0.052483	SNCB/LOXL4/MOXD1/P2RX4/LOX/S100A13/LOXL3/AOC3	8

downregulated.HGI_T1	GO:0001948	glycoprotein binding	11/574	102/16982	0.000659	0.056785	0.053593	CTSK/ADA2/GPNMB/CHRD/VWF/SDCBP/LDLR/FBXO6/CD4/B2M/CTSL	11
downregulated.HGI_T1	GO:0015037	peptide disulfide oxidoreductase activity	4/574	13/16982	0.000724	0.056785	0.053593	GSTO2/GSTO1/PDIA5/GLRX	4
downregulated.HGI_T1	GO:0016814	hydrolase activity, acting on carbon-nitrogen (but not peptide) bonds, in cyclic amidines	6/574	34/16982	0.000872	0.061528	0.05807	ADA2/APOBEC3D/APOBEC3G/AMPD3/APOBEC3F/MTHFD2	6
downregulated.HGI_T1	GO:0008009	chemokine activity	7/574	48/16982	0.001074	0.064968	0.061317	CCL5/CXCL11/CCL2/CCL7/CCL28/CCL26/CXCL16	7
downregulated.HGI_T1	GO:0043394	proteoglycan binding	6/574	36/16982	0.001192	0.064968	0.061317	CTSK/ADA2/GPNMB/CHRD/SDCBP/CTSL	6
downregulated.HGI_T1	GO:0004713	protein tyrosine kinase activity	15/574	181/16982	0.001268	0.064968	0.061317	FLT1/TIE1/ZAP70/FLT4/RET/EFNB3/MST1R/TEK/IL3RA/ALK/EIF2AK2/NRP2/CLK1/RIPK2/TXK	15
downregulated.HGI_T1	GO:0030547	receptor inhibitor activity	4/574	15/16982	0.00131	0.064968	0.061317	LYNX1/PCSK9/CCL5/ESR2	4
downregulated.HGI_T1	GO:0019955	cytokine binding	10/574	95/16982	0.00138	0.064968	0.061317	CSF3R/IL23R/ACVRL1/FZD4/CHRD/FAS/NRP2/IL12A/CD4/TWSG1	10
downregulated.HGI_T1	GO:0055102	lipase inhibitor activity	4/574	17/16982	0.002164	0.092831	0.087613	SNCB/APOC1/ANXA4/PINLYP	4
downregulated.HGI_T1	GO:0019239	deaminase activity	5/574	28/16982	0.002235	0.092831	0.087613	ADA2/APOBEC3G/AMPD3/APOBEC3F/GNPDA1	5
downregulated.HGI_T2	GO:0001948	glycoprotein binding	14/728	102/16982	0.000113	0.051359	0.048868	GPNMB/COL5A3/CTSK/CTSS/VWF/CD4/CHRD/SDCBP/CTSL/FBXO6/DMD/B2M/IDE/ERLEC1	14

downregulated.HGI_T2	GO:0004896	cytokine receptor activity	13/728	91/16982	0.00013	0.051359	0.048868	EBI3/IL23R/CSF3R/IL1RL2/IL22RA1/IL6R/IFNLR1/IL3RA/PRLR/CD4/IL1R1/IL7R/CCR10	13
downregulated.LGI_T1	GO:0005518	collagen binding	10/507	65/16982	2.11E-05	0.009062	0.008112	COL5A3/CTSK/PODN/COL14A1/MMP9/C1QTNF1/PCOLCE/ITGA10/CTSL/CTSB	10
downregulated.LGI_T1	GO:0019955	cytokine binding	12/507	95/16982	2.53E-05	0.009062	0.008112	IL23R/CCR1/CD74/CD36/CSF3R/ZFP36/NRP2/CD109/FAS/IL1R1/FZD4/NRP1	12
downregulated.LGI_T1	GO:0008201	heparin binding	15/507	152/16982	5.09E-05	0.012127	0.010855	CLEC3B/COL5A3/CCL2/GPNMB/FBLN7/CXCL11/APOE/SOD3/NRP2/PCOLCE/THBS3/PTN/VEGFA/SFRP1/NRP1	15
downregulated.LGI_T1	GO:0043394	proteoglycan binding	7/507	36/16982	7.98E-05	0.01418	0.012693	COL5A3/CTSK/GPNMB/SDCBP/CTSL/PTN/CTSB	7
downregulated.LGI_T1	GO:0005539	glycosaminoglycan binding	17/507	201/16982	0.000115	0.01418	0.012693	CLEC3B/COL5A3/BGN/TREM2/CCL2/GPNMB/FBLN7/CXCL11/APOE/SOD3/NRP2/PCOLCE/THBS3/PTN/VEGFA/SFRP1/NRP1	17
downregulated.LGI_T1	GO:0019838	growth factor binding	13/507	128/16982	0.000119	0.01418	0.012693	WISP2/CD36/WISP1/COL6A1/HTRA1/NRP2/CD109/NOV/DUSP1/PTN/IL1R1/S100A13/NRP1	13
downregulated.LGI_T1	GO:0004177	aminopeptidase activity	7/507	43/16982	0.000257	0.026278	0.023521	XPNPEP2/DPP4/CTSH/FAP/ANPEP/MMP14/ERAP2	7
downregulated.LGI_T1	GO:0008238	exopeptidase activity	11/507	106/16982	0.000325	0.029024	0.02598	XPNPEP2/DPP4/AEBP1/MME/CTSH/CPQ/FAP/TPP1/ANPEP/MMP14/ERAP2	11
downregulated.LGI_T1	GO:0015171	amino acid transmembrane transporter activity	9/507	80/16982	0.00062	0.049269	0.044101	SLC1A6/SLC6A16/SLC1A2/SLC7A8/SLC43A1/SLC6A9/SLC7A7/SLC38A7/SLC38A6	9

downregulate d.LGI_T1	GO:0030 547	receptor inhibitor activity	4/507	15/16982	0.000824	0.055636	0.0498	CCL5/LYNX1/PCSK9/ESR2	4
downregulate d.LGI_T1	GO:0001 948	glycoprotein binding	10/507	102/1698 2	0.00094	0.055636	0.0498	COL5A3/CTSK/GPNMB/BMPR1B/ SDCBP/MAP2/CTSL/PTN/BBC3/C TSB	10
downregulate d.LGI_T1	GO:1901 681	sulfur compound binding	16/507	221/1698 2	0.001014	0.055636	0.0498	CLEC3B/COL5A3/CCL2/GPNMB/F BLN7/CXCL11/APOE/SOD3/NRP2/ PCOLCE/THBS3/PTN/VEGFA/SFR P1/NRP1/DBI	16
downregulate d.LGI_T1	GO:0015 179	L-amino acid transmembrane transporter activity	7/507	54/16982	0.001066	0.055636	0.0498	SLC1A6/SLC1A2/SLC7A8/SLC43A 1/SLC6A9/SLC7A7/SLC38A7	7
downregulate d.LGI_T1	GO:0042 056	chemoattractant activity	5/507	27/16982	0.001089	0.055636	0.0498	CCL5/GPNMB/LGALS3/VEGFA/M IF	5
downregulate d.LGI_T1	GO:0033 218	amide binding	18/507	274/1698 2	0.001537	0.06739	0.060321	BACE2/CD74/PTGDR2/FOLR3/M ME/APOE/NPR1/HLA- DPA1/SLC7A8/TMEM158/TPP1/A NPEP/FTCD/COL4A3BP/PSAP/HL A-E/ERAP2/LDLRAD3	18
downregulate d.LGI_T1	GO:0004 896	cytokine receptor activity	9/507	91/16982	0.001571	0.06739	0.060321	GFRA2/IL23R/CCR1/CD74/EBI3/C SF3R/CCR10/IL3RA/IL1R1	9
downregulate d.LGI_T1	GO:0001 664	G-protein coupled receptor binding	17/507	254/1698 2	0.001686	0.06739	0.060321	CCL5/CCL2/APLN/RTP4/CXCL11/ RAMP1/AGT/CTHRC1/ASIP/SDCB P/FZD1/MAGI2/SFRP1/PSAP/PDE4 D/AGTR1/S1PR2	17
downregulate d.LGI_T1	GO:0005 496	steroid binding	9/507	92/16982	0.001697	0.06739	0.060321	ATP1A2/APOE/AKR1C2/AKR1C1/ ESR2/NPC1/RORA/NPC2/STARD4	9
downregulate d.LGI_T1	GO:0070 006	metalloaminopeptidas e activity	4/507	20/16982	0.0026	0.093231	0.083451	XPNPEP2/ANPEP/MMP14/ERAP2	4

downregulated.LGI_T1	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	12/507	157/16982	0.002676	0.093231	0.083451	CYP3A7/SCD/CYP27A1/FADS2/AKR1C2/AKR1C1/P4HA3/FADS1/MOXD1/POR/FMO5/SQLE	12
downregulated.LGI_T1	GO:0004879	RNA polymerase II transcription factor activity, ligand-activated sequence-specific DNA binding	6/507	48/16982	0.002913	0.093231	0.083451	NR4A2/NR4A3/NR4A1/ESR2/GLMP/RORA	6
downregulated.LGI_T1	GO:0098531	transcription factor activity, direct ligand regulated sequence-specific DNA binding	6/507	48/16982	0.002913	0.093231	0.083451	NR4A2/NR4A3/NR4A1/ESR2/GLMP/RORA	6
downregulated.LGI_T1	GO:0005319	lipid transporter activity	10/507	119/16982	0.002999	0.093231	0.083451	APOE/PLSCR1/PITPNC1/PLSCR4/FABP3/COL4A3BP/TNFAIP8L3/PLEKHA8P1/NPC1/STARD4	10
downregulated.LGI_T2	GO:0019955	cytokine binding	17/696	95/16982	2.76E-07	0.000216	0.000194	IL23R/CCR1/CSF3R/ACVRL1/CD74/CCR7/IL6R/TGFBR3/CD109/IL12A/FAS/ACKR4/NRP1/ITGA4/FZD4/IL1R1/LTBP1	17
downregulated.LGI_T2	GO:0008238	exopeptidase activity	15/696	106/16982	2.69E-05	0.007135	0.006417	XPNPEP2/DPP4/MME/ANPEP/CPQ/CTSH/TPP1/FAP/MMP14/ERAP2/LAP3/TRHDE/AGBL2/CPD/GGH	15
downregulated.LGI_T2	GO:0005539	glycosaminoglycan binding	22/696	201/16982	2.73E-05	0.007135	0.006417	CCL7/CCL2/CLEC3B/CXCL11/TREM2/SFRP1/JCHAIN/ACAN/PRELP/TGFBR3/SOD3/FGF7/TNFAIP6/COL13A1/RSPO3/APOE/NRP1/SLIT2/NAV2/CFH/FBLN7/APLP2	22

downregulated.LGI_T2	GO:0004177	aminopeptidase activity	9/696	43/16982	4.99E-05	0.008066	0.007254	XPNPEP2/DPP4/ANPEP/CTSH/FAP/MMP14/ERAP2/LAP3/TRHDE	9
downregulated.LGI_T2	GO:0008201	heparin binding	18/696	152/16982	5.15E-05	0.008066	0.007254	CCL7/CCL2/CLEC3B/CXCL11/SFRP1/PRELP/TGFBR3/SOD3/FGF7/COL13A1/RSPO3/APOE/NRP1/SLIT2/NAV2/CFH/FBLN7/APLP2	18
downregulated.LGI_T2	GO:0043394	proteoglycan binding	8/696	36/16982	8.35E-05	0.010892	0.009796	CTSK/CTSS/SDCBP/CTSL/SLIT2/FST/CFH/CTSB	8
downregulated.LGI_T2	GO:1901681	sulfur compound binding	22/696	221/16982	0.000114	0.012788	0.011502	CCL7/CCL2/CLEC3B/CXCL11/SFRP1/PRELP/TGFBR3/SOD3/FGF7/COL13A1/RSPO3/APOE/MGST1/NRP1/SLIT2/SOAT1/FST/NAV2/CFH/FBLN7/CD34/APLP2	22
downregulated.LGI_T2	GO:0033218	amide binding	25/696	274/16982	0.000163	0.015969	0.014362	FOLR3/CD74/PTGDR2/MME/PRLR/NPR1/HLA-DPA1/ANPEP/TMEM158/PPARG/APOE/TPP1/MGST1/SLC7A8/TAP1/PPIF/HLA-F/GSAP/ERAP2/HLA-E/SSTR2/TRHDE/PSAP/PRNP/COL4A3BP	25
downregulated.LGI_T2	GO:0005518	collagen binding	10/696	65/16982	0.000292	0.025389	0.022835	CTSK/CTSS/COL14A1/ITGA2/CTSL/C1QTNF1/SRGN/ITGA10/CTSB/DDR2	10
downregulated.LGI_T2	GO:0004896	cytokine receptor activity	12/696	91/16982	0.000332	0.025967	0.023354	IL23R/CCR1/CSF3R/EBI3/CD74/PRLR/CCR7/IL15RA/IL6R/ACKR4/IL7R/IL1R1	12

downregulated.LGI_T2	GO:0042277	peptide binding	22/696	242/16982	0.000415	0.029545	0.026572	CD74/PTGDR2/MME/PRLR/NPR1/HLA-DPA1/ANPEP/TMEM158/PPARG/APOE/TPP1/MGST1/SLC7A8/TAP1/PPIF/HLA-F/GSAP/ERAP2/HLA-E/SSTR2/TRHDE/PRNP	22
downregulated.LGI_T2	GO:0008235	metalloexopeptidase activity	9/696	57/16982	0.000478	0.031175	0.028038	XPNPEP2/ANPEP/CPQ/MMP14/ERAP2/LAP3/TRHDE/AGBL2/CPD	9
downregulated.LGI_T2	GO:0019838	growth factor binding	14/696	128/16982	0.000769	0.046314	0.041654	TEK/ACVRL1/IL6R/NOV/TGFBR3/CD109/NRP1/S100A13/DUSP1/IGFBP6/IL1R1/COL6A1/RHBDF2/LTBP1	14
downregulated.LGI_T2	GO:0070006	metalloaminopeptidase activity	5/696	20/16982	0.001058	0.059151	0.053199	XPNPEP2/ANPEP/MMP14/ERAP2/TRHDE	5
downregulated.LGI_T2	GO:0019199	transmembrane receptor protein kinase activity	10/696	80/16982	0.001549	0.080838	0.072704	TEK/ACVRL1/RET/TGFBR3/ALK/BMPR1B/NRP1/PDGFRL/DDR2/LTBP1	10
downregulated.LGI_T2	GO:0005507	copper ion binding	8/696	56/16982	0.001906	0.093174	0.083799	SNCB/SOD3/MOXD1/CP/LOXL4/S100A13/P2RX4/PRNP	8
downregulated.LGI_T2	GO:0016705	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	15/696	157/16982	0.002023	0.093174	0.083799	CYP3A7/AKR1C1/CH25H/AKR1C2/CYP27A1/PTGS1/TBXAS1/MOXD1/P4HA3/POR/SCD/ASPH/FMO5/CYP4V2/PHYH	15
upregulated.HGI_T1	GO:0030276	clathrin binding	8/417	60/16982	0.000103	0.037932	0.03379	SCLT1/SYT11/SYT16/DNER/SYTL5/SNAP91/SYT9/TRPC5	8
upregulated.HGI_T1	GO:0008373	sialyltransferase activity	5/417	21/16982	0.000128	0.037932	0.03379	ST8SIA4/ST8SIA6/ST6GALNAC5/ST8SIA2/ST6GALNAC3	5

upregulated. HGI_T1	GO:0008 201	heparin binding	12/417	152/1698 2	0.000378	0.056785	0.050585	CYR61/CTGF/THBS4/ADGRG1/A DAMTS5/THBS1/CRISPLD2/GRE M2/BMP4/NDNF/ADAMTS3/SELL	12
upregulated. HGI_T1	GO:1901 681	sulfur compound binding	15/417	221/1698 2	0.000384	0.056785	0.050585	CYR61/CTGF/THBS4/ADGRG1/A DAMTS5/THBS1/CRISPLD2/GRE M2/BMP4/OGDHL/RYR2/ACADL/ NDNF/ADAMTS3/SELL	15
upregulated. HGI_T1	GO:0005 230	extracellular ligand-gated ion channel activity	8/417	75/16982	0.000495	0.058639	0.052237	P2RX6/CHRNA5/GRIK2/GABRB1/ GRIN2B/GABRA4/GRID2/GABRA 2	8
upregulated. HGI_T1	GO:0030 594	neurotransmitter receptor activity	9/417	97/16982	0.000632	0.059489	0.052994	CHRNA5/HTR7/GRIK2/GABRB1/ GRIN2B/GABRA4/GRID2/GABRA 2/DRD1	9
upregulated. HGI_T1	GO:0008 013	beta-catenin binding	8/417	79/16982	0.000703	0.059489	0.052994	SOX9/CDH2/TCF7/SMAD7/LEF1/ GRIP1/SALL1/CTNND2	8
upregulated. HGI_T1	GO:0005 003	ephrin receptor activity	4/417	19/16982	0.001036	0.071675	0.06385	EPHB6/NTRK3/EPHB1/EPHA3	4
upregulated. HGI_T1	GO:0022 824	transmitter-gated ion channel activity	6/417	49/16982	0.001211	0.071675	0.06385	CHRNA5/GRIK2/GABRB1/GRIN2 B/GRID2/GABRA2	6
upregulated. HGI_T1	GO:0022 835	transmitter-gated channel activity	6/417	49/16982	0.001211	0.071675	0.06385	CHRNA5/GRIK2/GABRB1/GRIN2 B/GRID2/GABRA2	6
upregulated. HGI_T1	GO:0005 539	glycosaminoglycan binding	13/417	201/1698 2	0.001434	0.076525	0.068171	CYR61/CTGF/THBS4/ADGRG1/A DAMTS5/THBS1/CRISPLD2/GRE M2/BMP4/NDNF/ADAMTS3/SELL /SPOCK3	13
upregulated. HGI_T1	GO:0039 706	co-receptor binding	3/417	10/16982	0.001551	0.076525	0.068171	BMP4/BMP2/DKK2	3
upregulated. HGI_T1	GO:0016 917	GABA receptor activity	4/417	22/16982	0.001846	0.084041	0.074866	GPR156/GABRB1/GABRA4/GABR A2	4

upregulated. HGI_T1	GO:0004 222	metalloendopeptidase activity	9/417	115/1698 2	0.002116	0.089456	0.07969	MMP16/NLN/ADAMTS5/ADAMTS 16/ADAM22/ADAM20/ADAMTS20 /ADAMTS3/TLL1	9
upregulated. HGI_T1	GO:0015 276	ligand-gated ion channel activity	10/417	142/1698 2	0.002677	0.097833	0.087152	P2RX6/CHRNA5/GRIK2/GABRB1/ GRIN2B/RYR2/GABRA4/TRPM2/ GRID2/GABRA2	10
upregulated. HGI_T1	GO:0022 834	ligand-gated channel activity	10/417	142/1698 2	0.002677	0.097833	0.087152	P2RX6/CHRNA5/GRIK2/GABRB1/ GRIN2B/RYR2/GABRA4/TRPM2/ GRID2/GABRA2	10
upregulated. HGI_T1	GO:0003 777	microtubule motor activity	7/417	77/16982	0.002809	0.097833	0.087152	CENPE/DNAH5/KIF21A/DNAH11/ DNAH10/DYNC1I1/KIF5C	7
upregulated. HGI_T2	GO:0050 839	cell adhesion molecule binding	33/653	451/1698 2	0.000322	0.094845	0.090137	ITGA3/RPS2/FSCN1/EXOC3/CC2D 1A/TES/EEF2/CGN/FLNA/PPP1R1 3L/RANGAP1/EMD/EPHA2/PVR/D NAJB1/CNN2/CYR61/CTGF/EDIL3 /HSPA5/ADGRL1/MYH9/SLC9A3R 2/NRG1/PAK4/NLGN1/THBS4/AD AM22/THBS1/ADAMTS5/MPP7/P KP2/ANK3	33
upregulated. HGI_T2	GO:0000 982	transcription factor activity, RNA polymerase II core promoter proximal region sequence- specific binding	28/653	367/1698 2	0.000466	0.094845	0.090137	CC2D1A/SIX4/NR1D1/ATF4/TCF3/ SOX12/SSBP4/PITX2/EGR1/SSBP2 /JUND/RUNX2/GFI1/SOX9/SIX1/T OX2/OTX1/NR2E3/ONECUT2/TEF /FOXC2/FOXL2/HES1/HAND2/NK X3-2/LEF1/NKX6-1/SALL1	28
upregulated. HGI_T2	GO:0005 160	transforming growth factor beta receptor binding	8/653	49/16982	0.000512	0.094845	0.090137	SMAD6/SMAD7/BMP8A/RASL11B /GDF10/BMP4/BMP2/GDF6	8
upregulated. HGI_T2	GO:0008 083	growth factor activity	16/653	162/1698 2	0.00052	0.094845	0.090137	PDGFA/CD320/GFER/CTGF/NRG1 /NRG2/FGF18/BMP8A/THBS4/ERE G/GDF10/BMP4/BMP2/PDGFB/GD F6/IGF2	16

upregulated. HGI_T2	GO:0005 003	ephrin receptor activity	5/653	19/16982	0.000614	0.094845	0.090137	EPHB4/EPHA2/EPHB6/EPHB1/EP HA3	5
upregulated. LGI_T1	GO:0008 083	growth factor activity	13/378	162/1698 2	6.99E-05	0.039694	0.036487	FGF2/BMP8B/NRG1/PDGFA/CTGF /HBEGF/NRG2/NTF4/EREG/IGF2/ CXCL1/BMP4/GDF6	13
upregulated. LGI_T1	GO:0005 520	insulin-like growth factor binding	5/378	28/16982	0.000343	0.072418	0.066567	CYR61/CRIM1/ITGB4/IGFBP7/CT GF	5
upregulated. LGI_T1	GO:0005 125	cytokine activity	14/378	219/1698 2	0.00041	0.072418	0.066567	FGF2/BMP8B/CMTM8/EDN1/NRG 1/IL1A/TNFSF15/IL1B/TNFSF8/CC L20/CXCL1/BMP4/CXCL6/GDF6	14
upregulated. LGI_T1	GO:0008 201	heparin binding	11/378	152/1698 2	0.000605	0.072418	0.066567	FBN1/FGF2/CYR61/CTGF/NDNF/ HBEGF/TMEM184A/THBS1/ADA MTS3/BMP4/CXCL6	11
upregulated. LGI_T1	GO:0046 934	phosphatidylinositol-4,5-bisphosphate 3-kinase activity	7/378	66/16982	0.000645	0.072418	0.066567	FGF2/FGFR3/NRG1/PDGFA/HBEG F/NRG2/EREG	7
upregulated. LGI_T1	GO:0052 813	phosphatidylinositol bisphosphate kinase activity	7/378	69/16982	0.000844	0.072418	0.066567	FGF2/FGFR3/NRG1/PDGFA/HBEG F/NRG2/EREG	7
upregulated. LGI_T1	GO:0050 839	cell adhesion molecule binding	21/378	451/1698 2	0.001235	0.072418	0.066567	PSMB6/EPHA2/PDLIM5/FBN1/CB L/NOTCH3/DIAPH3/CIB2/ANLN/A DGRL1/CYR61/PPP1R13L/NRG1/E DIL3/CTGF/NLGN1/THBS1/ANK3/ PPME1/NRXN3/PKP2	21
upregulated. LGI_T1	GO:0035 004	phosphatidylinositol 3-kinase activity	7/378	74/16982	0.001281	0.072418	0.066567	FGF2/FGFR3/NRG1/PDGFA/HBEG F/NRG2/EREG	7

upregulated. LGI_T1	GO:1901 681	sulfur compound binding	13/378	221/1698 2	0.001402	0.072418	0.066567	FBN1/FGF2/CYR61/CTGF/NDNF/ HBEGF/TMEM184A/THBS1/ADA MTS3/BMP4/OGDHL/CXCL6/ACA DL	13
upregulated. LGI_T1	GO:0001 786	phosphatidylserine binding	5/378	38/16982	0.00146	0.072418	0.066567	SYT1/JPH2/THBS1/RASGRP1/MA RK1	5
upregulated. LGI_T1	GO:0005 088	Ras guanyl-nucleotide exchange factor activity	14/378	250/1698 2	0.001488	0.072418	0.066567	FGF2/FGFR3/NRG1/PDGFA/HBEG F/NRG2/FGD4/RASGRP1/ADRB1/ EREG/NEFL/FGD5/GFRA1/CSF2R B	14
upregulated. LGI_T1	GO:0070 412	R-SMAD binding	4/378	23/16982	0.00153	0.072418	0.066567	PMEPA1/MYOCD/PAX6/ANKRD1	4
upregulated. LGI_T1	GO:0003 779	actin binding	19/378	400/1698 2	0.001659	0.072498	0.06664	PAWR/PDLIM5/MYH11/WDR1/PA LLD/AFAP1/DIAPH3/ANLN/SHRO OM3/TPM1/FGD4/TAGLN3/MYO7 B/TAGLN/PPP1R9A/TRPC5/ANXA 8/LIMCH1/LCP1	19
upregulated. LGI_T1	GO:0005 539	glycosaminoglycan binding	12/378	201/1698 2	0.001868	0.075778	0.069655	FBN1/FGF2/CYR61/SUSD5/CTGF/ NDNF/HBEGF/TMEM184A/THBS1 /ADAMTS3/BMP4/CXCL6	12
upregulated. LGI_T1	GO:0019 838	growth factor binding	9/378	128/1698 2	0.002272	0.086042	0.07909	FGFR3/CYR61/CRIM1/ITGB4/PDG FA/IGFBP7/CTGF/THBS1/NGFR	9
upregulated. LGI_T1	GO:0004 713	protein tyrosine kinase activity	11/378	181/1698 2	0.002483	0.088146	0.081024	EPHA2/FGF2/FGFR3/CRIM1/NRG 1/HBEGF/STYK1/NRG2/EREG/EP HB1/CSF2RB	11

upregulated. LGI_T2	GO:0008 201	heparin binding	28/717	152/1698 2	3.91E-11	2.73E-08	2.52E-08	FGFRL1/CEL/RTN4R/THBS2/PTP RS/CTGF/CHRD/TMEM184A/APL P1/PTCH1/LIPI/MPIG6B/PGF/CCD C80/HBEGF/THBS1/ANOS1/ADA MTS15/NDNF/PCSK6/COL25A1/C OL5A1/CRISPLD2/POSTN/COMP/ BMP4/SELL/ADAMTS3	28
upregulated. LGI_T2	GO:0005 539	glycosaminoglycan binding	29/717	201/1698 2	6.85E-09	2.39E-06	2.20E-06	FGFRL1/CEL/RTN4R/THBS2/PTP RS/CTGF/CHRD/TMEM184A/APL P1/PTCH1/LIPI/MPIG6B/PGF/CCD C80/HBEGF/THBS1/ANOS1/ADA MTS15/NDNF/PCSK6/COL25A1/C OL5A1/CRISPLD2/POSTN/COMP/ BMP4/SELL/SPOCK3/ADAMTS3	29
upregulated. LGI_T2	GO:1901 681	sulfur compound binding	30/717	221/1698 2	1.59E-08	3.70E-06	3.41E-06	FGFRL1/CEL/RTN4R/THBS2/PTP RS/CTGF/CHRD/TMEM184A/APL P1/PTCH1/LIPI/MPIG6B/PGF/CCD C80/HBEGF/THBS1/ANOS1/ADA MTS15/NDNF/PCSK6/COL25A1/C OL5A1/CRISPLD2/POSTN/COMP/ BMP4/OGDHL/SELL/ACADL/AD AMTS3	30
upregulated. LGI_T2	GO:0005 201	extracellular matrix structural constituent	14/717	78/16982	4.20E-06	0.000732	0.000676	COL4A2/MUC6/COL11A1/COL1A 1/TUFT1/COL9A2/COL4A1/COL2 A1/ANOS1/COL9A3/COL5A1/CO MP/COL4A6/COL4A5	14
upregulated. LGI_T2	GO:0005 003	ephrin receptor activity	7/717	19/16982	7.50E-06	0.001046	0.000965	EPHB2/EFNB3/EFNA3/EPHB6/NT RK3/EPHB1/EPHA3	7
upregulated. LGI_T2	GO:0004 714	transmembrane receptor protein	12/717	64/16982	1.28E-05	0.001483	0.001368	FGFRL1/EPHB2/IGF1R/EFNB3/EF NA3/FGFR3/FLT4/ROR1/EPHB6/N TRK3/EPHB1/EPHA3	12

		tyrosine kinase activity							
upregulated. LGI_T2	GO:0048 407	platelet-derived growth factor binding	5/717	11/16982	4.94E-05	0.004922	0.004542	COL1A1/COL4A1/COL2A1/PDGF A/COL5A1	5
upregulated. LGI_T2	GO:0019 838	growth factor binding	16/717	128/1698 2	9.80E-05	0.008538	0.007879	FGFRL1/IGF1R/CTGF/COL1A1/FG FR3/COL4A1/COL2A1/FLT4/THBS 1/ITGB4/PCSK6/PDGFA/COL5A1/ NGFR/NTRK3/NKD2	16
upregulated. LGI_T2	GO:0019 199	transmembrane receptor protein kinase activity	12/717	80/16982	0.000126	0.009106	0.008403	FGFRL1/EPHB2/IGF1R/EFNB3/EF NA3/FGFR3/FLT4/ROR1/EPHB6/N TRK3/EPHB1/EPHA3	12
upregulated. LGI_T2	GO:0005 544	calcium-dependent phospholipid binding	9/717	47/16982	0.000131	0.009106	0.008403	PCLO/SYT15/C2CD4C/RPH3AL/D YSF/SYTL1/SYTL5/SYT12/ANXA 8	9
upregulated. LGI_T2	GO:0008 083	growth factor activity	18/717	162/1698 2	0.000171	0.010859	0.01002	CSPG5/FGF18/BMP1/CTGF/ARTN/ BMP8A/NRG1/PGF/HBEGF/BMP8 B/INHBB/PDGFA/GDF10/BMP4/N TF4/IGF2/GDF6/TGFA	18
upregulated. LGI_T2	GO:0004 713	protein tyrosine kinase activity	19/717	181/1698 2	0.000241	0.013987	0.012907	FGFRL1/EPHB2/IGF1R/FGF18/PR AG1/EFNB3/NRG1/TNK1/EFNA3/ MAP2K6/FGFR3/HBEGF/FLT4/RO R1/EPHB6/NTRK3/EPHB1/EPHA3/ CSF2RB	19

upregulated. LGI_T2	GO:0000 982	transcription factor activity, RNA polymerase II core promoter proximal region sequence-specific binding	30/717	367/1698 2	0.000441	0.02365	0.021823	E2F8/FOXC2/RUNX2/SKIL/TEF/F OXL2/ONECUT2/SOX4/SIX2/PITX 2/HOXA10/OTX1/KCNIP3/NPAS4/ TFAP2A/MEOX1/MYOCD/HEYL/ MAFB/DLX4/LEF1/HES1/GATA3/ BCL11A/NKX6-1/FOXA1/NKX3-2/ZIC1/NKX2-5/NFIB	30
upregulated. LGI_T2	GO:0005 160	transforming growth factor beta receptor binding	8/717	49/16982	0.000944	0.046987	0.043357	SMAD6/BMP8A/BMP8B/INHBB/S MAD7/GDF10/BMP4/GDF6	8
upregulated. LGI_T2	GO:0004 222	metalloendopeptidase activity	13/717	115/1698 2	0.001127	0.050856	0.046927	ADAMTS6/ADAMTS12/BMP1/MM P25/ADAMTS15/ADAM19/ADAM 22/ADAM28/ADAMTS20/ADAM20 /ADAMTS16/ADAMTS3/TLL1	13
upregulated. LGI_T2	GO:0001 228	transcriptional activator activity, RNA polymerase II transcription regulatory region sequence-specific binding	28/717	355/1698 2	0.001167	0.050856	0.046927	FOXC2/RUNX2/TEF/FOXL2/ONE CUT2/SOX4/CSRNP3/SIX2/PITX2/ HOXA10/GLIS1/OTX1/NPAS4/TF AP2A/MEOX1/MYOCD/TLX2/GA TA6/HEYL/MAFB/LEF1/GATA3/F OXA1/ALX1/ZIC1/HOXD8/NKX2-5/NFIB	28
upregulated. LGI_T2	GO:0001 077	transcriptional activator activity, RNA polymerase II core promoter proximal region sequence-specific binding	22/717	256/1698 2	0.001297	0.053178	0.04907	FOXC2/RUNX2/TEF/FOXL2/ONE CUT2/SOX4/SIX2/PITX2/HOXA10/ OTX1/NPAS4/TFAP2A/MEOX1/M YOCD/HEYL/MAFB/LEF1/GATA3 /FOXA1/ZIC1/NKX2-5/NFIB	22
upregulated. LGI_T2	GO:0008 373	sialyltransferase activity	5/717	21/16982	0.001531	0.059302	0.054721	ST3GAL5/ST8SIA2/ST6GALNAC3/ ST6GALNAC5/ST6GAL2	5

upregulated. LGI_T2	GO:0070 696	transmembrane receptor protein serine/threonine kinase binding	4/717	13/16982	0.00166	0.060901	0.056196	SMAD6/BMP8A/BMP8B/BMP4	4
upregulated. LGI_T2	GO:0005 088	Ras guanyl-nucleotide exchange factor activity	21/717	250/1698 2	0.002184	0.076097	0.070219	PLEKHG3/FGF18/KALRN/ARTN/E PS8L2/FGD4/NRG1/SPTBN2/GRIN 2D/FGFR3/HBEGF/SPTB/PLEKHG 4B/ARHGEF16/PDGFA/RASGRP1/ NEFL/GRIN2B/FGD5/GFRA1/CSF2 RB	21
upregulated. LGI_T2	GO:0070 412	R-SMAD binding	5/717	23/16982	0.002361	0.078378	0.072324	SMAD6/MYOCD/ANKRD1/PAX6/ PMEPA1	5
upregulated. LGI_T2	GO:0033 612	receptor serine/threonine kinase binding	4/717	15/16982	0.002963	0.090045	0.083089	SMAD6/BMP8A/BMP8B/BMP4	4
upregulated. LGI_T2	GO:0000 987	core promoter proximal region sequence-specific DNA binding	26/717	343/1698 2	0.002971	0.090045	0.083089	E2F8/PER1/SMAD6/RUNX2/SKIL/ FOXL2/ONECUT2/SIX2/PITX2/HO XA10/OTX1/KCNIP3/NPAS4/TFAP 2A/MEOX1/MYOCD/MAFB/DLX4/ LEF1/GATA3/BCL11A/NKX6-1/NKX3-2/ZIC1/NKX2-5/NFIB	26
upregulated. LGI_T2	GO:0001 159	core promoter proximal region DNA binding	26/717	346/1698 2	0.003339	0.096968	0.089478	E2F8/PER1/SMAD6/RUNX2/SKIL/ FOXL2/ONECUT2/SIX2/PITX2/HO XA10/OTX1/KCNIP3/NPAS4/TFAP 2A/MEOX1/MYOCD/MAFB/DLX4/ LEF1/GATA3/BCL11A/NKX6-1/NKX3-2/ZIC1/NKX2-5/NFIB	26

Comparative GO enrichment analysis of DE genes in subgroups at T1 and T2. Illustrates comparison of GO terms in four comparison HGI_T1, HGI_T2, LGI_T1, LGI_T2. Log2 fold change >1, FDR <0.05.

APPENDIX 6.5 Genes related to NHEJ pathways

alt.EJ pathway		
ENSEMBL Gene ID	**Gene symbol**	**Reference**
ENSG00000005156	LIG3	Audebert *et al.*, 2004
ENSG00000020922	MRE11	Makharashvili *et al.*, 2014
ENSG00000051341	POLQ	Wood and Doublie, 2016
ENSG00000073050	XRCC1	Audebert *et al.*, 2004
ENSG00000101773	RBBP8	Jazayeri et al., 2006
ENSG00000101773	MIR4741	Audebert *et al.*, 2004
ENSG00000104320	NBN	Makharashvili *et al.*, 2014
ENSG00000113522	RAD50	Makharashvili *et al.*, 2014
ENSG00000129484	PARP2	Fouquin et al., 2017
ENSG00000143799	PARP1	Audebert *et al.*, 2004
ENSG00000168496	FEN1	Wu et al. 1999

c-NHEJ pathway		
ENSEMBL Gene ID	**Gene symbol**	**Reference**
ENSG00000012048	BRCA1	Bau et al., 2006
ENSG00000042088	TDP1	
ENSG00000067369	TP53BP1	
ENSG00000079246	XRCC5	
ENSG00000080345	RIF1	Bhargava et al., 2018
ENSG00000111802	TDP2	
ENSG00000137337	MDC1	
ENSG00000149311	ATM	
ENSG00000152422	XRCC4	Callebaut et al., 2006
ENSG00000152457	DCLRE1C	
ENSG00000157212	PAXIP1	
ENSG00000166169	POLL	Moon et al., 2014
ENSG00000174405	LIG4	Callebaut et al., 2006
ENSG00000187736	NHEJ1	
ENSG00000196419	XRCC6	
ENSG00000253729	PRKDC	Meek et al., 2008
ENSG00000123358	NR4A1	Malewicz et al., 2011

APPENDIX 6.6 Metrics analysis of whole exome sequence data.

READ_MATRICS	Parental	TCL1	TCL1_EV	CP_05	CP_15	CP_16	CP_21	CP_38	CP_105
PF_READS_TOTAL	44570746	50419946	45036228	51984350	61617928	41704240	43775940	53287952	43589054
PF_READS_ALIGNED	44528318	50363962	44992646	51886160	61553809	41650877	43712988	53215212	43535046
PF_READS_ALIGNED (%)	99.90	99.89	99.90	99.81	99.90	99.87	99.86	99.86	99.88
PF_HQ_ALIGNED_READS	42633158	48398293	43228704	49754025	59144667	40043134	41987519	51179568	41873467
PF_HQ_ALIGNED_READS (%)	95.74	96.10	96.08	95.89	96.09	96.14	96.05	96.17	96.18
PF_READS_ALIGNED_IN_PAIRS	44499800	50327348	44965590	51869182	61509822	41614810	43695888	53192442	43496118
PF_READS_ALIGNED_IN_PAIRS (%)	99.936	99.9273	99.9399	99.9673	99.9285	99.9134	99.9609	99.9572	99.9106
PF_READS_PCR DUPLICATES	4006608	4543308	4073873	4911840	5620222	3676471	4049825	5347968	3868634
PF_READS_PCR DUPLICATES (%)	8.99	9.01	9.04	9.44	9.12	8.81	9.24	10.03	8.87
PF_READS_UNALIGNED	42428	55984	43582	98190	64119	53363	62952	72740	54008
PF_READS_UNALIGNED (%)	0.10	0.11	0.10	0.19	0.10	0.13	0.14	0.14	0.12
PF_READS_CHIMERAS (%)	0.39	0.55	0.48	0.46	0.51	0.33	0.47	0.41	0.36
PF_READS_ADAPTER (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PF_READS_ALIGNED_TO '+' (%)	49.53	49.85	49.65	49.56	49.58	49.46	49.48	49.13	49.66
PF_READS_ALIGNED_TO '-' (%)	50.44	50.12	50.33	50.41	50.39	50.51	50.49	50.84	50.31
MEAN_READ_LENGTH	148.99	149.18	149.21	148.42	149.22	149.20	148.44	148.40	149.18
PF_ALIGNED_BASES	6610913204	7484868723	6690408508	7669168839	9151586401	6191244642	6461617795	7865358398	6470236752

PF_HQ_ALIGNED_Q20_BASES	6200595336	7029604915	6294102187	7149764745	8595948137	5813759023	6021390347	7332367360	6078012804
PF_HQ_ALIGNED_Q20_BASES (%)	97.89	97.65	97.84	97.12	97.68	97.60	96.92	96.84	97.59
PF_MISMATCH_RATE	0.003601	0.003816	0.003475	0.004199	0.003815	0.003906	0.004394	0.004471	0.003926
PF_HQ_ERROR_RATE	0.003513	0.003726	0.00339	0.004114	0.003725	0.003818	0.004305	0.004378	0.003833
PF_INDEL_RATE	0.000116	0.000125	0.000123	0.000124	0.000127	0.000119	0.000128	0.00013	0.000122

Metrics analysis of whole exome sequence data. The metrics analysis was carried out using metric using Picard tools. Picard terms: PF reads= reads that passing illumina's filter "chastity filter", PF_HQ_ALIGNED_READS = the number of PF reads that were aligned to the reference sequence with a mapping quality of Q20 or higher. PF_HQ_ALIGNED_Q20_BASES= the subset of PF_HQ_ALIGNED_BASES where the base call quality was Q20 or higher. PF_MISMATCH_RATE = the rate of bases mismatching the reference for all bases aligned to the reference sequence. PF_HQ_ERROR_RATE= the fraction of bases that mismatch the reference in PF HQ aligned reads. PF_INDEL_RATE= the number of insertion and deletion events per 100 aligned bases.

APPENDIX 6.7 List of chromosomal segments with LOH in pre-crisis and post-crisis samples.

Sample ID	Chrom	Start	End	Estimated fraction of cells with LOH for the segment	Estimated total copy number	Estimated minor allele copy number
TCL1	8	163635	43152208	0.815716	1	0
TCL1	11	87042	51412083	0.574105	1	0
TCL1	18	47692	15271065	0.635019	1	0
TCL1	19	110812	22940046	0.482675	1	0
TCL1_EV	8	163635	43152208	0.7371425	1	0
TCL1_EV	9	70176945	141071670	0.5262767	1	0
TCL1_EV	11	87042	51459173	0.5262767	1	0
TCL1_EV	18	47692	14999981	0.5262767	1	0
TCL1_EV	19	110812	19843354	0.5262767	1	0
TCL1_EV	19	58991723	58991723	0.7371425	1	0
CP_05	2	53977948	91843433	0.529979	1	0
CP_05	8	48068797	118847703	0.320168	1	0
CP_05	11	87142	50003801	0.529979	2	0
CP_05	11	77774953	134257527	0.529979	1	0
CP_05	17	25628903	40055020	0.529979	2	0
CP_05	17	51900375	81052262	0.529979	2	0
CP_05	18	47624	15271365	0.529979	1	0
CP_05	X	200817	75397506	0.529979	2	0
CP_05	X	75397764	110035322	0.529979	2	0
CP_05	X	110366424	119388696	0.529979	2	0
CP_05	X	1.19E+08	155254945	0.529979	2	0
CP_15	2	53992546	89459357	0.355435	1	0
CP_15	2	99438330	242842568	0.712868	2	0
CP_15	11	87142	36615604	0.712868	1	0
CP_15	18	47624	694268	0.257435	1	0
CP_15	18	13645590	15271365	0.257435	1	0
CP_15	19	21281210	30102937	0.712868	2	0
CP_16	7	146741075	158937176	0.536508	1	0
CP_16	8	42977762	132051958	0.536508	1	0
CP_16	11	210042	36472789	0.536508	2	0
CP_16	16	943130	31613054	0.536508	2	0
CP_16	18	47624	18531256	0.536508	1	0
CP_16	X	200817	105193696	0.536508	2	0
CP_16	X	105197095	154774756	0.536508	2	0
CP_21	8	48068797	66539641	0.702911	2	0
CP_21	8	66617096	146278811	0.911985	2	0
CP_21	10	43128255	135473081	0.539248	2	0

CP_21	11	87142	36615604	0.702911	2	0
CP_21	16	66962	32077867	0.302562	3	0
CP_21	17	21536113	81052262	0.619344	3	0
CP_21	19	110712	23675183	0.905954	3	0
CP_38	8	121970500	126114686	0.677845	1	0
CP_38	11	55111500	60620585	0.93919	2	0
CP_38	11	89537560	116629815	0.93919	2	0
CP_38	11	121970500	134257527	0.93919	1	0
CP_38	16	34681636	90162604	0.93491	2	
CP_38	18	2891432	15271365	0.860961	2	0
CP_38	21	47611093	48084329	0.939190	2	0
CP_105	1	14907	2702442	0.908688	2	0
CP_105	8	93074835	136657378	0.906206	2	0
CP_105	11	49942114	63176367	0.896661	2	0
CP_105	11	82443417	101375490	0.891905	2	0
CP_105	11	108536095	113295255	0.896661	2	0
CP_105	11	124805895	126873778	0.896661	2	0
CP_105	16	249198	31486059	0.908688	2	0
CP_105	16	52473299	88931953	0.908688	2	0
CP_105	17	11459232	16526379	0.908688	2	0
CP_105	17	17394810	18088423	0.908688	2	0
CP_105	19	425131	3612341	0.902592	3	0
CP_105	19	22363847	24310803	0.902592	3	0

LOH was counted when the minor allele CNV =0.

APPENDIX 6.8 Annotating of genes that associated with segmental amplification in CP_05 and CP_38

CP_05

Gene symbol	Ensembl Gene ID	Chrom	Stand	Start	End
OR4F21	ENSG00000176269	chr8	-	116085	117024
RPL23AP53	ENSG00000223508	chr8	-	158344	182318
ZNF596	ENSG00000172748	chr8	+	182199	197339
FBXO25	ENSG00000147364	chr8	+	356807	419875
AK056623	ENSG00000272005	chr8	+	424454	427966
TDRP	ENSG00000180190	chr8	-	439789	494845
C8orf42	ENSG00000180190	chr8	-	494664	494998
ERICH1	ENSG00000104714	chr8	-	614199	681226
ERICH1-AS1	ENSG00000237647	chr8	+	687586	1087777
LOC286083	ENSG00000253640	chr8	-	1244293	1250827
DLGAP2	ENSG00000198010	chr8	+	1449531	1656642
LOC440518	ENSG00000269332	chr19	+	22779058	22786352
BC030765	ENSG00000268038	chr19	-	22798358	22806773
ZNF492	ENSG00000229676	chr19	+	22817125	22850472
ZNF99	ENSG00000213973	chr19	-	22934984	22966973

CP_38

Gene symbol	Ensembl Gene ID	Chrom	Stand	Start	End
ZNF626	ENSG00000188171	chr19	-	20802744	20844402
ZNF66	ENSG00000160229	chr19	+	20959109	20977403
ZNF85	ENSG00000105750	chr19	+	21106058	21126783
ZNF430	ENSG00000118620	chr19	+	21203425	21242852
ZNF714	ENSG00000160352	chr19	+	21264952	21301256
ZNF431	ENSG00000196705	chr19	+	21324839	21368805
ZNF708	ENSG00000182141	chr19	-	21473962	21512212
ZNF738	ENSG00000172687	chr19	+	21541734	21562104
ZNF493	ENSG00000196268	chr19	+	21579920	21591601
CR627135	ENSG00000268119	chr19	-	21634855	21646686
LINC00664	ENSG00000268658	chr19	+	21666175	21682696
ZNF429	ENSG00000197013	chr19	+	21688436	21721079
BC033373	ENSG00000268240	chr19	-	21770233	21776808
ZNF100	ENSG00000197020	chr19	-	21906842	21933585
LOC641367	ENSG00000269845	chr19	+	21933546	21936240
ZNF43	ENSG00000198521	chr19	-	21987750	22019010
ZNF208	ENSG00000160321	chr19	-	22115759	22193745
ZNF257	ENSG00000197134	chr19	+	22235265	22273903
ZNF676	ENSG00000196109	chr19	-	22361902	22379753
ZNF729	ENSG00000196350	chr19	+	22469251	22499978

ZNF98	ENSG00000197360	chr19	-	22573898	22605148
LOC440518	ENSG00000269332	chr19	+	22779058	22786352
LINC01785	ENSG00000268038	chr19	-	22798358	22806773
ZNF492	ENSG00000229676	chr19	+	22817125	22850472
ZNF99	ENSG00000213973	chr19	-	22934984	22966973
ZNF728	ENSG00000269067	chr19	-	23157684	23170205
BC037873	ENSG00000261615	chr19	-	23237624	23244067
ZNF730	ENSG00000183850	chr19	+	23299776	23330017
ZNF724P	ENSG00000196081	chr19	-	23406606	23433172
AK023040	ENSG00000213971	chr19	-	23441499	23457032
BC038574	ENSG00000167232	chr19	-	23521417	23541617
AK022793	ENSG00000269416	chr19	-	23582035	23586295
ZNF675	ENSG00000197372	chr19	-	23835707	23870017
ZNF681	ENSG00000196172	chr19	-	23921996	23941693
RPSAP58	ENSG00000205246	chr19	+	23945815	24010919
ZNF726	ENSG00000213967	chr19	+	24097683	24116769
AK125686	ENSG00000269289	chr19	-	24101882	24140732
AK092080	ENSG00000268362	chr19	+	24216246	24249490
ZNF254	ENSG00000213096	chr19	+	24216246	24312654

Annotating of genes that associated with segmental amplification in in CP_05 and CP_38. Overlapped genes coloured in orange.