Identification and Characterisation of Telomere Regulatory and Signalling Pathways after Induction of Telomere Dysfunction

Author: John Revie

Thesis submitted to the University of Glasgow in fulfilment of the requirements for the degree of Doctor of Philosophy

School of Life Sciences
College of Medical, Veterinary & Life Sciences
Institute of Cancer Sciences
Wolfson Wohl Cancer Research Centre
University of Glasgow

May 2016

Abstract

Telomeres are DNA-protein complexes which cap the ends of eukaryotic linear chromosomes. In normal somatic cells telomeres shorten and become dysfunctional during ageing due to the DNA end replication problem. This leads to activation of signalling pathways that lead to cellular senescence and apoptosis. However, cancer cells typically bypass this barrier to immortalisation in order to proliferate indefinitely. Therefore enhancing our understanding of telomere dysfunction and pathways involved in regulation of the process is essential. However, the pathways involved are highly complex and involve interaction between a wide range of biological processes. Therefore understanding how telomerase dysfunction is regulated is a challenging task and requires a systems biology approach. In this study I have developed a novel methodology for visualisation and analysis of gene lists focusing on the network level rather than individual or small lists of genes. Application of this methodology to an expression data set and a gene methylation data set allowed me to enhance my understanding of the biology underlying a senescence inducing drug and the process of immortalisation respectively. I then used the methodology to compare the effect of genetic background on induction of telomere uncapping. Telomere uncapping was induced in HCT116 WT, $p21^{-/-}$ and $p53^{-/-}$ cells using a viral vector expressing a mutant variant of hTR, the telomerase RNA template. $p21^{-/-}$ cells showed enhanced sensitivity to telomere uncapping. Analysis of a candidate pathway, Mismatch Repair, revealed a role for the process in response to telomere uncapping and that induction of the pathway was p21 dependent. The methodology was then applied to analysis of the telomerase inhibitor GRN163L and synergistic effects of hypoglycaemia with this drug. HCT116 cells were resistant to GRN163L treatment. However, under hypoglycaemic conditions the dose required for ablation of telomerase activity was reduced significantly and telomere shortening was enhanced. Overall this new methodology has allowed our group and collaborators to identify new biology and improve our understanding of processes regulating telomere dysfunction.

List of Tables

List of Figures

Table of Contents

Abbreviations

ALT	Alternative lengthening of telomere
ATM	Ataxia telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3 related
BLM	Bloom syndrome protein
CDK	Cylin-dependent kinase
ChIP	Chromatin Immunoprecipitation
CIMP	CpG island methylator phenotype
CST	CTC1, STN1 and TEN1 complex
CTE	C-terminal extension
DDR	DNA damage response
DKC	Dyskeratosis Congenita
DNA	Deoxyribonucleic acid
DNA-SCARS	DNA segments with chromatin alterations reinforcing senescence
GEO	Gene Expression Omnibus
HCT116	Homo sapiens colon colorectal carcinoma
HDR	Homology directed repair
HHS	Hoyeraal-Hreidarsson syndrome
ITS	Interstitial telomeric sequences
MEF	Mouse embryonic fibroblasts
NHEJ	Non-homologous end joining
OIS	Oncogene induced senescence
PCTG	Polycomb target genes
PML	Promyelocytic Leukemia
PNO	Process network enrichment analysis methodology
RNA	Ribonucleic Acid
ROS	Reactive oxygen species
RT	Reverse transcriptase domain
SAHF	Senescence associated heterochromatin foci
SASP	Senescence associated secretory phenotype
TEN	telomerase N-terminal
TRBD	TR-binding domain
WT	Wild Type

Acknowledgments

I would like to thank my supervisor Prof W. Nicol Keith for offering me the opportunity to undertake a PhD in his laboratory, for the excellent guidance provided and allowing me the freedom to take on such a project. I would also like to thank my advisor Dr Alan Bilsland for his continued support and guidance throughout the project, his technical advice and support providing great ideas over the course of the project.

Thanks also to all of the group's members, past and present, for their support during my time at the Wolfson Wohl Cancer Research Centre. Special thanks go to Dr Claire Cairney and Sharon Burns for their supervision and support in the lab.

Finally I'd like to thank my very patient family for their support in this endeavour.

This work was supported by Cancer Research UK and I am grateful to this funding body for making this opportunity possible.

Declaration

I declare that, except where explicit reference is made to the contribution of others, that this thesis is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution.

Signature:

Printed name: John Revie

Chapter 1: Introduction

1.1. Telomere structure and function

As DNA replicates the telomere shortens with each replication cycle due to the end replication problem (Harley et al., 1990). In order to counter this problem mammals have evolved a nucleoprotein structure at the ends of their chromosomes, called telomeres, which prevent this shortening from reaching regions of DNA vital to cell survival and also to prevent recognition of the telomeres as DNA damage by endogenous proteins (de Lange, 2010). There are two main components to the telomere: the telomeric DNA repeat structure which is a series of TTAGGG repeats generated by the Telomerase enzyme and the multi-subunit protein complex which protects it called Shelterin.

1.1.1. Telomeric DNA

The DNA at chromosome ends in human cells consists of 5–15 kb of telomeric TTAGGG repeat. The length of the telomeres varies between species. For example mice telomere can range from 25-40kb (Blackburn, 2001; Palm and de Lange, 2008). The majority of these repeats are double-stranded but the very end of each chromosome is made up of a single-stranded 3′ overhang which can vary from 50-500bp in mammals. This single stranded sequence is key in formation of a t-loop structure where telomere ends can fold back on themselves (Makarov et al., 1997; Palm and de Lange, 2008). The size of the loops appears to be variable as loops ranging from 1 – 25kb have been observed. However, end processing seems to be a controlled process where the 5’ and 3’ sequences may be differentially regulated. In human telomeres the 5’ end is homogeneous, where most telomeres have the sequence AATCCCAATC-5’, whilst the 3’ end can be more variable (Sfeir et al., 2005). The overall length of the telomeric DNA is also regulated and the protein structure which coats the telomeric DNA plays a role in this process. Telomeres are regulated by a negative feedback loop which blocks the action of telomerase. As telomeres become too long the telomerase pathway is inhibited. When telomeres are short this control is relaxed and telomerase may restore its length. This seems to be

regulated by the amount of Shelterin complex bound to the telomeres, which increases as telomeres become longer and have a greater probability of inhibiting telomerase (Smogorzewska and de Lange, 2004).

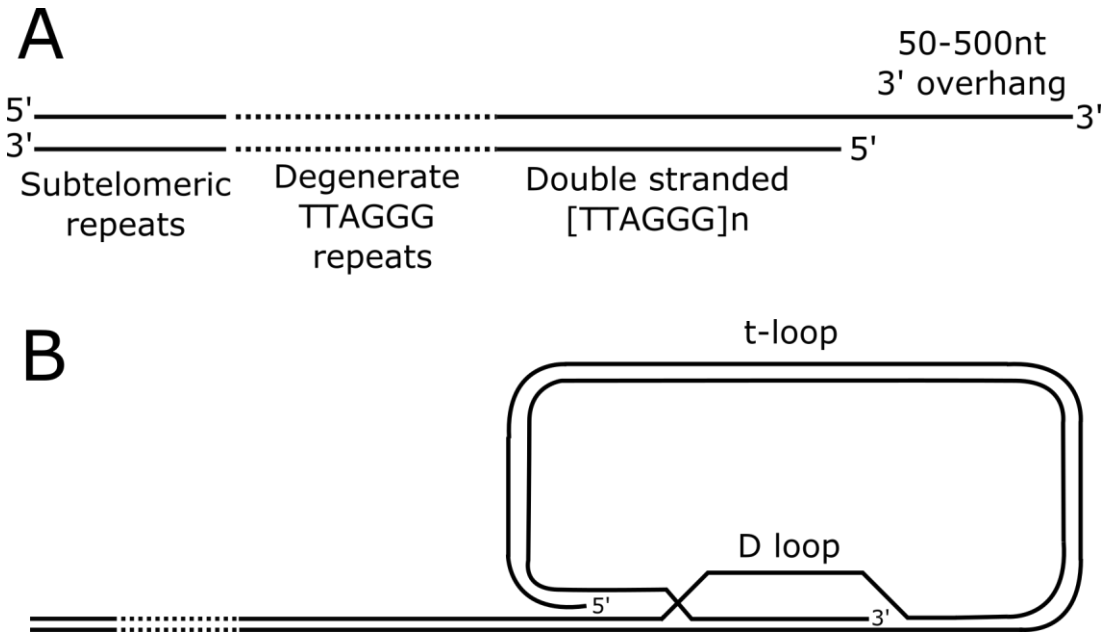


Figure 1.1. The structure of human telomeric DNA.

(A) Human chromosomes end in an array of TTAGGG repeats that varies in length. Proximal to the telomeric repeats is a segment of degenerate repeats and subtelomeric repetitive elements. The telomere terminus contains a long G-strand overhang. (B) Schematic of the t-loop structure, a lariat-like configuration that arises by strand invasion of the telomeric 3' G-overhang into the upstream telomeric double-stranded DNA forming an internal D-loop. The size of the loop is variable.

In addition to the T-loop formation, the high G content of the telomeres enables G-quadruplexes to form in the single stranded regions. The G-quadruplex structures play an important regulatory role at the telomeres and are implicated in telomere protection, suppression of recombination and inhibition of telomerase extension (Lipps and Rhodes, 2009). They have been shown to have an inhibitory effect on telomere elongation in vitro (Zahler et al., 1991). Supporting this is evidence that G-quadruplex formation at the 3' end of telomeres prevents telomerase accessibility to the telomere (Tang et al., 2008). These G-quad structures also have inhibitory roles against other enzymes, such as helicases BLM and other RecQ family helicases (Liu et al., 2010). Unravelling of these structures is also important. It is thought that end processing enzymes

such as telomerase require a free single stranded 3' telomere end. It was demonstrated that telomere G-quadruplex unwinding, telomerase extension of the telomere and alternative lengthening of telomere (ALT) mechanism need a 3' tail of 6, 8 and 12nt respectively (Wang et al., 2011b). Formation of G-quadruplexes also appears to be directly regulated. It has been shown that the yeast telomerase subunit Estp1 is able to form G-quadruplex structures in single stranded telomeric G-rich DNA (Tong et al., 2011; Zhang et al., 2010). Further to this Est1p mutants lead to disrupted G-quadruplex formation, telomere shortening and cellular senescence suggesting that G-quadruplexes are essential in telomere protection (Zhang et al., 2010). Overall, it is understood that these G-quadruplex structures are important in the regulation of multiple enzymes at the telomeres thereby regulating telomere length and integrity and may be an interesting drug target for ageing and cancer therapies. It also suggests that G-quadruplexes may play an important role in regulation of telomere dysfunction detection and signalling within the cell.

1.1.2. Shelterin and accessory proteins

Shelterin is made up of 6 core subunits: TRF1, TRF2, POT1, TIN2, TPP1 and Rap1. TRF1 and TRF2 bind the double strand repeat sequence and provide specificity to the telomeres. POT1 binds single strand repeats and interacts with TRF2 (Yang et al., 2005). Interconnecting these proteins are TIN2, TPP1 and Rap1. TRF2 and POT1 are essential in hiding the telomeric DNA from detection by ATM and ATR. TRF2 facilitates the formation of the t-loop structure which protects the double stranded component of telomeres from detection by ATM whilst POT1 protects against detection by ATR by outcompeting RPA in the single strand region (de Lange, 2010). There are also a number of non Shelterin proteins with important roles at the telomeres. Traditionally, the Shelterin proteins have been distinguished by a number of criteria. They are most abundant at telomeres without accumulating elsewhere. Their primary functions are specific to telomeres and they are present at the telomeres throughout the cell cycle (de Lange, 2005). While recent findings that many of the Shelterin proteins have extra-telomeric functions bring this definition of Shelterin into question, non-Shelterin proteins fail to meet some or all of these criteria. Attempts to find

additional key Shelterin components using mass-spectroscopy techniques have yet to be successful. This suggests that we have a fairly complete view of the essential proteins in the Shelterin complex (Liu et al., 2004; O'Connor et al., 2004; Ye et al., 2004).

The core Shelterin proteins act as a platform allowing for recruitment of various proteins from a diverse range of pathways. The complexity of regulatory pathways involved in telomere maintenance was highlighted by a study looking at interaction partners of the core Shelterin proteins where over 300 proteins were identified (Lee et al., 2011). While some accessory proteins and complexes have known roles, most of the accessory proteins identified require further characterisation. Some of these proteins also appear to have roles essential to telomere function. For example, BLM helicases have been implicated in t-loop formation (Opresko et al., 2004). Another example is the CST complex which is essential in telomere replication and regulation of telomerase activity at the telomeres (Wu et al., 2012). CST is composed of three RPA-like proteins, CTC1, STN1 and TEN1 and has been shown to be a terminator of telomerase activity. It does this by binding the 3' single stranded telomeric DNA and displacing telomerase when the overhang reaches a specific length (Chen and Lingner, 2013; Chen et al., 2012a).

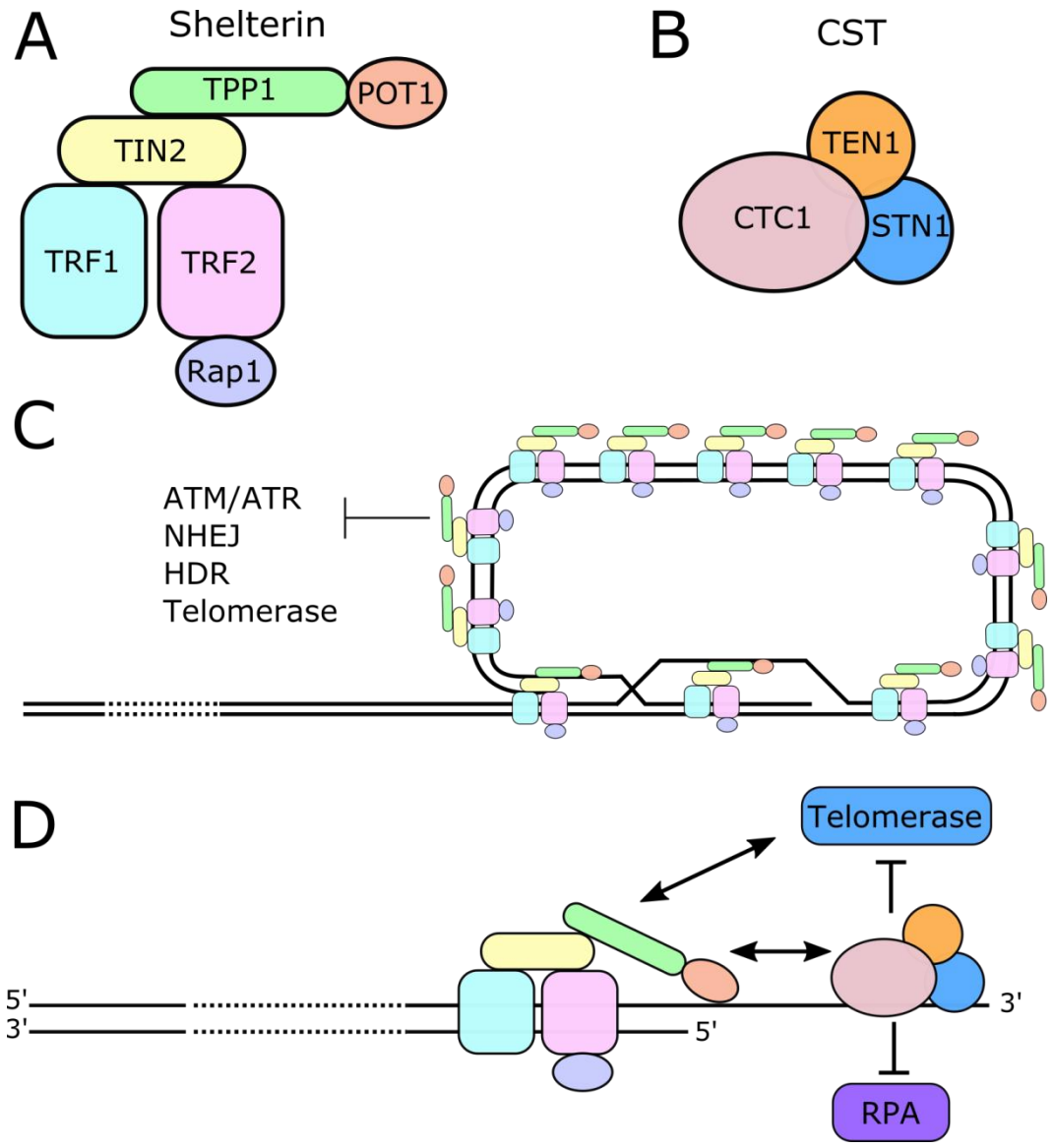


Figure 1.2. The Core Shelterin Subunits and CST complex at telomeres.

(A) The six-member Shelterin complex and its specific interactions. (B) The end-binding CST trimeric complex. (C) Telomeres in a closed T-loop configuration. Most Shelterin components are found interacting with double-stranded telomeric DNA and TPP1/POT1 that are found associated with the 3' G-overhang and D-loop. This configuration inhibits the access of telomerase, the checkpoint kinases ATM and ATR to telomeres and prevents NHEJ and HDR at telomeres. (D) The orchestrated actions of Shelterin and CST to regulate telomerase access, which promotes telomere elongation. At the telomere ends, telomerase can be recruited by the TPP1/POT1 interaction. The telomere extension by telomerase is terminated when the CST complex binds to the newly synthesized, single-stranded 3' G-overhang (protruding 3' single-strand). The CST complex inhibits the binding of RPA to telomeres due to CST higher affinity for single-stranded G-rich DNA.

There has been much work done to study the canonical role of the Shelterin proteins and their function at the telomeres. However, there is also increasing evidence of extra-telomeric functions for many of the Shelterin components. While we may have a strong understanding of their function at telomeres our knowledge of their role elsewhere in the cell is still incomplete. TIN2, TPP1 and POT1 have all been shown to localize in the cytoplasm (Chen et al., 2007). TIN2 also has mitochondrial localisation signals which can target it to the mitochondria and RNAi knockdown of TIN2 led to enhanced oxygen consumption and ATP synthesis, suggesting a role for the protein in metabolic regulation (Chen et al., 2012b). TRF2 has also been implicated in regulation of neuronal genes influencing neural tumour and stem cell differentiation (Zhang et al., 2008). In addition, TRF2 has been found to play a role in homologous recombination (HR) outside of the telomere (Bradshaw et al., 2005; Mao et al., 2007). Deregulation of a number of genes was also associated with RAP1 deletion in mice, preferentially affecting genes in subtelomeric regions (Martinez et al., 2010). Furthermore, RAP1 was shown to complex with IKKs (IκB kinases) in the cytoplasm and regulate NF-κB activation thereby affecting regulation of genes targeted by NF-κB (Teo et al., 2010). Genome-wide ChIP-seq identified a range of interstitial telomeric sequences (ITSs) to which TRF2 and RAP1 can bind (Yang et al., 2011). This finding for TRF2 was confirmed in another study which also found TRF1 to have a role at these sites (Simonet et al., 2011). As some of these sites are proximal to genes this may implicate RAP1, TRF1 and TRF2 in regulation of transcription. Overall, a number of non-telomeric functions for most of the Shelterin proteins have been uncovered indicating that they may play a wider role in regulation of biological processes than previously thought. Further study is needed to uncover possible roles for the Shelterin protein that may occur as a result of being displaced from the telomere during telomere dysfunction.

1.1.3. Chromatin structure at the telomeres

Telomeres themselves also have a specific chromatin structure and this has been found to be important in regulation of telomeric length. Whilst the telomere itself is coated in the Shelterin complex, nucleosomes are also present (Lejnine

et al., 1995). The subtelomeric regions also contain methylated DNA and H3K9 and H3K20 methylated nucleosomes. Loss of methyltransferase activity in these regions has been associated with abnormally long telomeres in mouse models null for Suv39h1 and Suv39h2 (Garcia-Cao et al., 2004). Sirt6 H3K9 deacetylase was also found to target the telomere and depletion of Sirt6 leads to telomere dysfunction and subsequent senescent phenotype. These findings suggest that telomere length and structure are also regulated by subtelomeric chromatin structure (Michishita et al., 2008).

Interactions between Shelterin and telomeric nucleosomes are not well understood, although recent findings suggest TRF2 binding increases nucleosomal spacing in a cell-cycle dependent fashion (Galati et al., 2012). A recent study investigated changes in histone biosynthesis as a result of replicative senescence. O'Sullivan et al found that human fibroblasts treated with bleomycin or cultured to replicative senescence demonstrated p53 independent downregulation of histones H3 and H4 (O'Sullivan et al., 2010). They propose that this removes the capacity of proliferating cells to restore their original chromatin structure, post replication, leading to a build-up of abnormal chromatin which reinforces the DDR signal eventually leading to senescence. Notably, expression of telomerase was sufficient to prevent accumulation of global chromatin reorganisation indicating that chromatin effects initiated at telomeres are capable of amplification throughout the nucleus.

1.1.4. Telomerase

Telomerase is an RNA containing reverse transcriptase made up of two main components, hTERT (catalytic component) and hTR (RNA component), both of which are required for telomerase activity and both have highly complex regulatory networks (Bilsland et al., 2009; Lafferty-Whyte et al., 2009). There are also the accessory proteins dyskerin, NOP10, NHP2 and GAR1 which when mutated can lead to a number of telomerase deficiency diseases such as Dyskeratosis Congenita (DKC). The main function of Telomerase is to add telomeric repeats to the end of chromosomes and is recruited by Shelterin (Blackburn and Collins, 2011; Nandakumar and Cech, 2013; Palm and de Lange,

2008). Once at the telomere the single stranded region can be utilised by Telomerase as a primer which binds to the RNA template (Cech, 2004). Synthesis of telomeric can then begin and continue, following repositioning of the 3' end (Greider, 1991).

hTR is the RNA component of the telomerase enzyme. Although the size and sequence is not highly conserved between species the four main functional features are. The four functional features are: the template for reverse transcription, the pseudoknot domain, a stem-loop which interacts with TERT and a 3' element required for RNA stability (Theimer and Feigon, 2006). The human TR (hTR) is 451nt long and contains all of the features mentioned (Zhang et al., 2011). The 3' stabilizing element is an H/ACA domain made up of two hairpins connected by a short single-stranded stretch, the H-box, and a terminal ACA region (Egan and Collins, 2012; Mitchell et al., 1999). The cofactors dyskerin, NOP10, NHP2 and GAR1 associate with the H/ACA domain to form the H/ACA complex which is essential for accumulation, enzymatic activity and correct localisation of telomerase (Kiss et al., 2010). The importance of the H/ACA complex is highlighted by telomerase deficiency diseases associated with mutations in these factors (Armanios and Blackburn, 2012).

The hTERT protein contains four key functional domains:- the telomerase N-terminal (TEN) domain; TR-binding domain (TRBD); reverse transcriptase domain (RT) and C-terminal extension (CTE) (Blackburn and Collins, 2011). The TEN domain is involved in recruitment of telomerase to the telomeres and in catalysis of telomeric repeat synthesis (Jurczyluk et al., 2011; Robart and Collins, 2011; Schmidt et al., 2014; Wu and Collins, 2014). The TRBD and reverse transcriptase domain are involved in association with hTR (Lai et al., 2001). The active site of the enzyme is made up by the reverse transcriptase domain and CTE. The reverse transcriptase domain of hTERT demonstrates homology with the reverse transcriptase domains of retrotransposon and retroviral reverse transcriptases (Lingner et al., 1997; Nakamura et al., 1997).

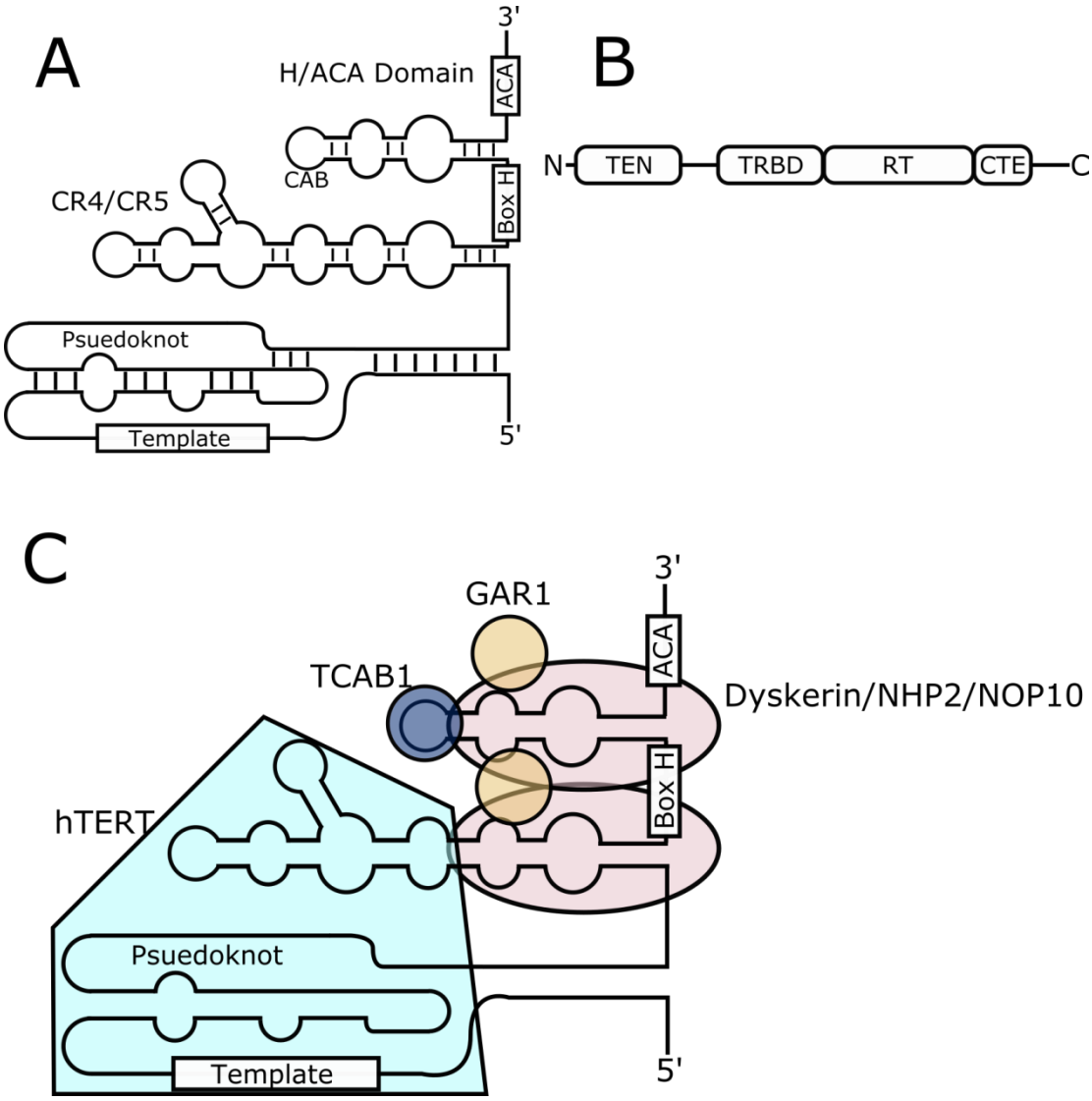


Figure 1.3. Telomerase and associated subunits.

(A) Secondary structure of the hTR. The pseudoknot and CR4/CR5 domains interact with hTERT. The template for telomeric repeat synthesis is indicated. Two copies of the H/ACA complex, each composed of dyskerin, NHP2, NOP10, and GAR1, associate with the H/ACA domain of hTR to stabilize the RNA in the nucleus.TCAB1 interacts with the CAB box to facilitate telomerase localization to Cajal bodies. (B) hTERT contains four functional domains. The telomerase N-terminal (TEN) domain participates in catalysis and drives telomerase localization to telomeres. The TR-binding domain (TRBD) interacts with hTR. The reverse transcriptase (RT) and C-terminal extension (CTE) form the catalytic core of telomerase. (C) Fully assembled and catalytically active telomerase enzyme.

Telomerase is highly regulated and implicated in oncogenesis and other diseases. For example, the hTERT promoter is controlled by a complex transcriptional network (Bilsland et al., 2009). Recently promoter mutations have been identified which affect expression levels of hTERT in cancers, highlighting the importance of telomerase in oncogenesis (Horn et al., 2013; Huang et al., 2013). They have also have been shown to be the most frequent mutation in some cancer types (Heidenreich et al., 2014). These mutations were associated with increased hTERT expression, telomerase activity and telomere length (Borah et al., 2015). While increased telomerase activity can be oncogenic, deficiencies in telomerase activity can also lead to diseases such as Dyskeratosis Congenita (DKC) (Armanios and Blackburn, 2012).

1.1.5. Telomere Structure and Cancer

Linear eukaryotic chromosomes are complex structures which pose multiple issues to cells that need to be regulated in order to maintain proliferative capacity and genome integrity. Due to the end replication problem, telomeres shorten with each replication cycle which, over time, can lead to loss of genetic information (Harley et al., 1990). This progressive shortening can also lead to Shelterin becoming displaced from the telomeres, resulting in the telomeric DNA becoming exposed to endogenous DNA damage recognition proteins. In normal cells, this leads to activation of DNA damage signalling response through induction of tumor suppressors p53 and p16, eventually causing senescence or apoptosis (d'Adda di Fagagna, 2008). This is an important tumour suppressor mechanism in normal cells as it prevents the growth of potentially cancerous cells. However, it is also a double edged sword and can be a source of genomic instability leading to oncogenesis (Shay and Wright, 2011). As telomere ends become exposed they can form chromosomal fusions which can then lead to breakage-fusion-bridge cycles in cells that escape senescence or apoptosis (van Steensel et al., 1998).

In cancer cells telomere structure can appear relatively normal whilst being regulated very differently. For example, while normal cells experience progressive shortening of the telomeres, cancer cells often demonstrate

maintenance of telomere length due to reactivation of telomerase (Degerman et al., 2010). As a result, the telomere structure is maintained, telomeres do not reach critical length and therefore do not activate DNA damage response associated with short telomeres, allowing indefinite proliferation. Alternatively, cancers can also demonstrate highly unstable telomeres, where the normal structure is compromised, but still proliferate (Chin et al., 2004; Rudolph et al., 2001). This is typically due to some form of interference in the downstream signalling pathways that usually lead to cellular arrest in normal cells. For example, it has been shown in yeast that by severing the DNA damage signalling pathway, cells can continue to divide despite having dysfunctional telomeres (Carneiro et al., 2010).

1.2. Causes and outcome of telomere dysfunction

Telomeres are usually bound and protected from recognition as DNA damage by Shelterin (de Lange, 2010). However, critically short telomeres fail to recruit functional Shelterin complexes, becoming exposed and initiating a DNA damage response which in turn leads to activation of various downstream signalling processes, leading to cellular arrest (senescence) or apoptosis. In normal cells, telomeres progressively shorten, due in part to the end-replication problem, eventually leading to replicative senescence, which can be accelerated by oxidative damage and other events (Cairney et al., 2012). This shortening can be blocked by the telomerase enzyme which extends telomeres by adding TTAGGG repeats to the ends of DNA (Collins and Mitchell, 2002). In normal tissue homeostasis telomerase activity is restricted to highly proliferating tissues (Wright et al., 1996). However cancer cells that have escaped senescence typically demonstrate reactivation of telomerase (Shay and Wright, 2011).

The reverse transcriptase and RNA components of telomerase (hTERT and hTR) are both required for telomerase activity and both have highly complex regulatory networks (Bilsland et al., 2009; Lafferty-Whyte et al., 2010b). For example, inhibition or mutation of hTR has been shown to induce telomere shortening and decrease telomerase activity (Feng et al., 1995). Telomeres can also be elongated via the alternative lengthening of telomeres (ALT) pathway,

which lengthens telomeres via a HR (Homologous recombination) mechanism (Brault and Autexier, 2011; Lafferty-Whyte et al., 2009). However, this mechanism is suppressed in most cells by proteins such as TRF2 (Celli et al., 2006).

1.2.1. DNA Damage Response and the Cell Cycle

The focus here will be on telomere shortening induced senescence, also known as replicative senescence However, stimuli which induce senescence are still being discovered and the mechanisms involved are extensively characterised (Campisi and d'Adda di Fagagna, 2007; Collado et al., 2007; Salama et al., 2014). Senescence is a form of irreversible proliferative arrest in which cells exit the cell cycle, entering G0, and is typically associated with prolonged arrest of the cell cycle with a G1 DNA content (Buttitta and Edgar, 2007). Senescence inducing stimuli signal through multiple pathways, often activating p53, which all essentially converge on the activation of cycling-dependent kinase (CDK) inhibitors p21, p15, p21 and p27. CDK-cyclin complex inhibition results in arrest of proliferation and the crucial senescence inducing component is hypo-phosphorylated Rb (Chicas et al., 2010).

Telomeres essentially act as molecular clocks keeping track of the replicative history of primary cells (Harley et al., 1990). Telomeres shorten during consecutive cell divisions and eventually reach critical length where the telomeric structure becomes compromised. This results in a type of senescence known as replicative senescence. As telomeres become critically short they are sensed by cells as a type of DNA damage and trigger a DNA-damage response (DDR). This DDR is mediated by ATM, ATR, CHK1 and CHK2 which phosphorylate and activate cell cycle proteins such as p53. This leads to expression of p21 which can inhibit CDK-cyclin complexes such as those involving CDK2 (Campisi and d'Adda di Fagagna, 2007). Replicative senescence has also been linked to two crucial tumour suppressors, p16 and ARF, which are encoded by the CDKN2A locus (also known as INK4A and ARF). p16 has been shown to be an inhibitor of CDK4 and CDK6 whilst ARF regulated p53 stability through inactivation of E3

ubiquitin protein ligase MDM2 which degrades p53 (Gil and Peters, 2006; Kim and Sharpless, 2006).

Senescence is a key barrier to cancer progression. Oncogenic signalling flux increases during the early stages of tumorigenesis until it reaches a threshold that activates the key tumour suppression pathways p16 and p53. This leads to inhibition of the cell cycle, causing cells to enter senescence, which prevents the expansion of precancerous cells. Indeed, senescence is detectable in benign tumours (Collado et al., 2007) and it has been demonstrated that cells which undergo damage-induced senescence can be removed by immune-mediated clearance (Hoenicke and Zender, 2012; Kang et al., 2011; Xue et al., 2007). This is due to a combination of direct detection of senescent cells by T helper cells (Kang et al., 2011) and by recruitment of inflammatory phagocytic cells (Xue et al., 2007), which are probably attracted by senescence associated secretory protein (SASP) factors. (Campisi, 2013; Kuilman and Peeper, 2009). The effectiveness of pro-senescent therapies has also been demonstrated in malignant tumours. It has been shown that activation of p53 in hepatocellular carcinomas and sarcomas induces senescence. This is then followed by tumour cell elimination (Ventura et al., 2007; Xue et al., 2007).

Over the past decade research has demonstrated that senescence can play both beneficial and detrimental roles (Munoz-Espin and Serrano, 2014). Transient induction of senescence can lead to the elimination of damaged cells. However, persistent senescence or the inability to clear senescent cells can lead to detrimental effects. The general purpose of senescence is to clear unwanted cells which can provide a similar role as apoptosis which is also important in elimination of damaged cells. Both of these mechanisms are particularly relevant in cancer and ageing: processes which are both associated with the accumulation of severe cellular damage. Therefore, senescence is an essential barrier to cancer progression and to senescent cells which accumulate with ageing.

Various stresses on the telomere eventually lead to degradation of the telomere structure leaving the telomeric DNA exposed. Exposure of the telomere ends can

be tumorigenic as telomeres can be highly fusogenic leading to chromosome end to end fusions and thus further damage to genomic DNA (van Steensel et al., 1998). However critically shortened or uncapped telomeres are also recognised as DNA damage by the processes outlined above and DDR foci have been shown to co-localise with telomeric DNA in senescent cells when telomeres become sufficiently shortened. These “telomere dysfunction induced foci” have been found to contain a variety of DDR proteins such as ATM and γH2AX (d'Adda di Fagagna et al., 2003; Takai et al., 2003).

Signalling of these pathways varies depending on the method of induction. For example, it was recently proposed that oncogene induced senescence (OIS) can cause hyper-replication and accumulation of DNA damage leading to a robust DDR via that ATR pathway (Di Micco et al., 2006) whereas replicative senescence typically operates via the ATM pathway (Herbig et al., 2004). Similarly, the checkpoint at which arrest is enforced appears to vary and can occur at various points within the cell cycle. Recently TRF2/POT1 inhibition was shown to cause arrest at the G2/M checkpoint via the ATM/ATR pathway (Thanasoula et al., 2012a). An earlier review also suggested that checkpoint activation could vary depending on the severity or type of damage response where the G1/S checkpoint may be more sensitive to exposed DNA ends than the G2/M checkpoint (d'Adda di Fagagna, 2008). It has also been demonstrated in yeast that telomere detection can be avoided by breaking the signalling pathway through prevention of the chromatin alterations required for stable recruitment of damage signalling factors (Carneiro et al., 2010).

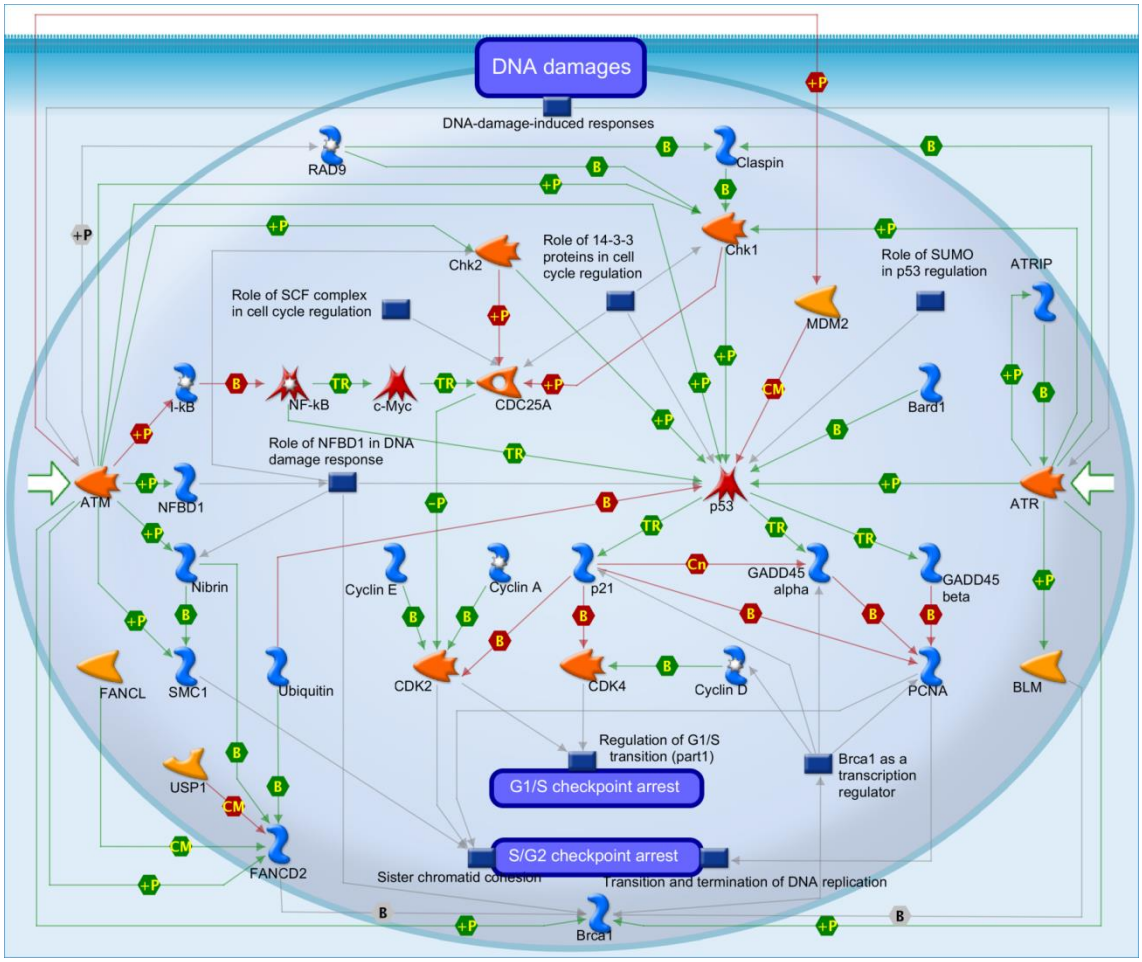


Figure 1.4. ATM/ATR regulation of the G1/S checkpoint.

Green, red, or gray arrows represent positive, negative, or unspecified effects, respectively. Arrow labels denote interaction types where B = binding, Cn = competition, TR = transcriptional regulation, P = phosphorylation, and CM = covalent modification.

1.2.2. Chromatin regulation

The senescent phenotype involves changes in many biological processes and widespread gene expression changes so it is unsurprising that chromatin remodelling plays a role. Substantial remodelling occurs around key regulators in a variety of senescence pathways (Dimauro and David, 2009). Accumulation of genomic damage and specific areas of hypermethylation have been linked with ageing and may be a precursor to cellular senescence (DePinho, 2000; Oakes et al., 2003). Development of heterochromatin, as directed by pRb, has also been associated with induction of senescence as key proliferative genes become

suppressed such as E2F target genes which promote cell-cycle entry and proliferation (Narita et al., 2003).

A number of specific histone remodelling activities play roles in senescence. For example, downregulation of histone methyltransferase EZH2 during senescence causes de-repression of the INK4A locus containing the promoter of the senescence inducer p16 (Bracken et al., 2007). Other histone demethylases, such as Jhdm1b, have also been found to prevent senescence via downregulation of the p53/pRb pathway (He et al., 2008). Similarly, global changes in acetylation are associated with senescence, with inhibition of histone deacetylases with sodium butyrate and trichostatin A, leading to senescent phenotype in primary human fibroblasts (Ogryzko et al., 1996).

Senescence associated heterochromatin foci (SAHF) have been proposed as a new senescence marker and occur as a result of telomere dysfunction. These are manifest as DAPI foci each containing a single tightly condensed chromosome, co-localised with various chaperones and heterochromatin marks including heterochromatin proteins 1, H3K9me3, HMGA1 and A2 and the macroH2A histone variants, but excluding H3K4me3 and H3K9Ac (Narita et al., 2003; Zhang et al., 2007; Zhang et al., 2005). SAHF were originally proposed to assemble specifically on E2F dependent promoters, dependent on pRb and HIRA/anti-silencing factor (ASF1a)/histone H3, thereby silencing key cell-cycle regulators (Narita et al., 2003; Zhang et al., 2007; Zhang et al., 2005). Recently, histone demethylase Jarid1b was also shown to associate with several E2F dependent cell cycle promoters in a pRb dependent manner in senescent mouse embryonic fibroblasts (MEFs), which correlates with decreased expression and loss of H3K4-me3 (Nijwening et al., 2011). Although SAHF are not readily formed in MEFs (Kennedy et al., 2010), overexpression of Jarid1a in human fibroblasts was sufficient to result in pRb dependent silencing of E2F target genes, along with global loss of H3K4Me3, cell cycle arrest, and induction of senescence markers including SAHF (Chicas et al., 2012).

However, a recent study examining assembly of SAHF in 3 normal cell strains under a range of senescence induction triggers concluded that they are not

required for senescence and form in a cell type and trigger dependent fashion, correlated with induction of p16. In particular, Ras induced senescence appears to promote robust SAHF responses in a range of cell types whereas other triggers showed cell-specific results and SAHF were not detected in tumour sections (Kosar et al., 2011). Similar results were observed by the d'Adda di Fagagna group. However, that group also observed that heterochromatin is not a *de facto* barrier to proliferation, since cells induced to bypass oncogene induced senescence by ATM or p53 knockdown retain SAHF-like heterochromatin, though lose H3K9Me3 marks at E2F target genes and fail to suppress these. Interestingly, SAHF may also act to constrain damage signalling, thereby promoting cell longevity during senescence (Di Micco et al., 2011).

1.2.3. Senescence associated secretory phenotype

Senescent cells have also been found to display a secretory signature known as Senescence associated secretory phenotype (SASP) which has been found to be irreversible once established (Coppe et al., 2010). SASP reinforces the phenotype of senescent cells but paradoxically can be oncogenic to normally proliferating cells (Young and Narita, 2009). It was recently reported that treatment of fibroblasts with histone deacetylase inhibitors promotes SASP in the absence of DNA damage, indicating the likelihood of a close link to chromatin regulation in senescent cells (Pazolli et al., 2012). Moreover, co-inoculation of treated fibroblasts with luciferase expressing HaCaT cells in NOD/SCID mice promoted tumour formation. While the SASP has been shown to be a double edged sword, promoting oncogenesis whilst reinforcing senescence, there have been other indicators of the anti-tumour activity of SASP due to the link it has to clearance of senescent cells *in vivo* via induction of inflammatory cytokines which activate the innate immune response (Xue et al., 2007). p16 induced senescence does not produce a normal SASP whereas p53 has been found to restrain the SASP and is thought to be a tumour suppressor mechanism via suppression of the pro-tumorigenic/pro-inflammatory microenvironment which the SASP creates (Coppe et al., 2008).

However, it was reported in other studies that persistent DNA damage is required for SASP, with temporal alterations evident in the composition of damage signalling foci in senescent cells. Sites of active repair following mild DNA damage are rapidly resolved, while more severe damage gives rise to persistent foci apparently lacking DNA synthesis but containing active p53 and Chk2, stabilised by H2AX and associated with nuclear Promyelocytic Leukemia (PML) bodies. These have been termed DNA-SCARS (DNA segments with chromatin alterations reinforcing senescence) (Rodier et al., 2009; Rodier et al., 2011). DNA-SCARS were required both for the growth arrest in senescent cells and for delayed secretion of IL-6 and IL-8, although viral oncogene expression could decouple growth arrest.

A number of soluble factors are associated with SASP such as interleukins, inflammatory cytokines and growth factors which mediate a diverse range of signalling activities. IL-6 is an example of a key SASP molecule which has been found to reinforce the senescent phenotype whilst promoting tumourigenesis in nearby cells suggesting that the oncogenic effect of SASP is an unintended side effect (Young and Narita, 2009). One initiated SASP is self-reinforcing which leads to irreversible expression of the phenotype. For example, CXCR-2 binding chemokines were found to be expressed as a result of OIS resulting in a self-amplifying secretory network which reinforces growth arrest (Acosta et al., 2008). Extracellular proteases are also linked to SASP via regulation of other SASP proteins. A number of MMPs (matrix metalloproteinases) were upregulated in human fibroblasts undergoing stress-induced premature senescence and may regulate the SASP via direct interaction with other SASP molecules such as chemokines (Liu and Hornsby, 2007; Van Den Steen et al., 2003).

SASP has also been associated with telomere dysfunction and replicative senescence. As mentioned, persistent DNA damage signalling has been shown to lead to SASP. Telomere dysfunction induction of persistent DDR is a major cause of cellular senescence (d'Adda di Fagagna et al., 2003). Senescent cells activate production of reactive oxygen species (ROS) and secrete SASP associated pro-inflammatory peptides (Coppe et al., 2010; Passos et al., 2010). In senescent fibroblasts and in oncogene-induced senescence, SASP is closely controlled

through signalling by NF-κß (Chien et al., 2011; Freund et al., 2011). Both senescence-associated ROS and NF-κß driven pro-inflammatory cytokines, such as IL-6 and IL-8, contribute to feedback loops that stabilise oncogene or stress-induced senescence (Acosta et al., 2008; Kuilman et al., 2008). SASP has been shown to be a self-reinforcing senescence mechanism, however exactly how SASP may feedback to enhance telomere dysfunction is not well understood. Recently, Jurk et al demonstrated that low level chronic inflammation induced in a mouse model led to premature ageing. They found that telomere-dysfunctional senescent cells accumulated in nfkb1$^{-/-}$ cells and that this accumulation could be blocked by anti-inflammatory or antioxidant treatment. This was shown to be caused by ROS-mediated exacerbation of telomere dysfunction leading to cellular senescence.

While SASP in the context of cancer has not been specifically linked to induction of telomere dysfunction, several SMS factors are linked with ageing and replicative senescence. For example, IL-6 (Hong et al., 2007), GRO1 (Himi et al., 1997), PAI1 (Goldstein et al., 1994) and TGFß (Carrieri et al., 2004) all increase with age and appear to be involved in induction of cellular senescence. This suggests that SASP is also associated with the ageing process in an autocrine, paracrine and possibly endocrine fashion. As these increase with age, they are implicated in a role in telomere dysfunction and replicative senescence. However, the exact feedback mechanisms by which SASP may feedback directly to the telomeres is not well understood but may be important in the wider context of regulation of telomere dysfunction.

1.2.4. Pathways involved are complex and a systems view is required for understanding

These genes involved in regulation of the interconnected subsystems outlined above constitute a substantial part of the senectome and typical response to telomere dysfunction, although other processes are also involved. Glycolytic flux appears to play a role in regulation of senescence in MEFs, likely by influencing cellular ROS levels (Kondoh et al., 2005). Widespread cytoskeletal changes are also involved (Wang and Jang, 2009). All these findings point to the complexity

of telomere dysfunction induced signalling and a requirement for diverse sub-systems to cooperate in senescence decision making. Hence, although sentinels such as p53 and pRb clearly are critical, global signal integration is likely to be at least as important in determining the overall phenotype. Indeed, in some cases core effectors seem able to be bypassed. Deficiency of Skp2, a component of the SCF-Skp2 E3 ubiquitin ligase, co-operates with deficiency of Pten to promote senescence independently of p53 and DNA damage (Lin et al., 2010).

Furthermore, although senescent cells are arrested, ongoing signalling still is a key feature. Computational analysis of the expression of DNA damage or secretory senescence genes in public data sets, demonstrated that a DNA damage signalling spike precedes the onset of increased inflammatory signalling in mesenchymal stem cell senescence (Lafferty-Whyte et al., 2010a). The Campisi group also reported that senescence associated secretion of inflammatory factors IL-6 and IL-8 follows delayed kinetics and requires the onset of persistent DNA damage signalling in several model systems (Rodier et al., 2009; Rodier et al., 2011). Other temporally regulated adaptations in senescence include transient co-localisation of HP1 proteins and the histone chaperone HIRA with PML bodies preceding SAHF formation (Zhang et al., 2007; Zhang et al., 2005), paracrine effects of the secretory response itself and widespread changes in metabolism such as increased autophagy (Singh et al., 2012).

Thus, the senectome is complex but also is dynamically regulated, suggesting the need for a high level of coordination. A recent proteomics study of ERK pathway dynamics in PC12 cell differentiation concluded that cell fate determination in that system relies on distributed control rather than a master switch (von Kriegsheim et al., 2009). The possible benefits of distributed control in irreversible decision making include temporal control of commitment, noise reduction, and a high degree of signal integration. These are likely to be key requirements for many other cell fate decisions presumably including senescence (Bar-Yam et al., 2009). Indeed, the existence of stochastic senescence inducer signals in normal cells, at least in culture, suggests a substantial requirement for noise filtration and/or gating to ensure faithful senescence responses only when

the severity of insult is commensurate with irreversible arrest (Lawless et al., 2012; Martin-Ruiz et al., 2004; Passos et al., 2007).

1.3. Why is understanding Telomere Dysfunction signalling important?

Although first described over 50 years ago (Hayflick and Moorhead, 1961), there is substantially renewed recent interest in mechanistic regulation of senescence as a result of recognition that tumour cells also undergo senescence-like arrest responses to various genotoxic stimuli including chemotherapeutic agents ("accelerated senescence") (Cairney et al., 2012). Hence, there is considerable interest in the potential for developing senescence-targeted cancer therapeutics. This broad goal implies the need for a systems-level view of the regulation of telomere dysfunction. A number of markers have been linked with senescence, which occurs as a result of telomere dysfunction, such as elevated p21 and p16 levels, the senescence associated secretory phenotype (SASP), senescence associated beta-galactosidase (SA-βGal) staining, senescence associated heterochromatin foci (SAHF) and changes in morphology where senescent cells become flat and enlarged *in vitro*. Beyond these relatively well defined markers however, it is also clear that senescence involves an extremely broad range of biological processes from telomere homeostasis, DNA damage and inflammatory signalling to chromatin regulation and metabolism (Capparelli et al., 2012; Rai and Adams, 2012). Though some key pathways have been characterised and provide interesting targets for drug discovery, the complexity of their interactions is not well defined and neither, as yet, is the complete set of gene targets able to elicit or modulate accelerated senescence responses. Thus, target discovery remains a goal to be framed not only in context of known sentinels such as the core p53/pRb pathways, but with a view to the global regulation of telomere dysfunction.

1.3.1. Oncogenesis, Replicative Immortality and Cancer progression

In the normal setting, telomere dysfunction is caused by telomere attrition (Harley et al., 1990) which typically leads to senescence or apoptosis and is one of the key barriers which limits the proliferation potential of cancerous cells (Prieur and Peeper, 2008). However, most tumours are reliant on the ability to proliferate indefinitely and have managed to overcome the barrier of telomere dysfunction (Durant, 2012). This is thought to occur mainly via two mechanisms: maintenance of the telomere itself or by blocking and/or rerouting the signalling pathway. One method is the re-expression of the telomerase subunits, hTERT and hTR, leading to maintenance or reconstitution of telomere length (Degerman et al., 2010). Increased telomerase activity has been demonstrated in a range of cancers resulting in maintenance of telomere length, therefore bypassing telomere attrition leading to immortality (Shay and Wright, 1996). Maintenance of telomere length can also occur via activation of the ALT mechanism, alternate lengthening of telomeres, which maintains telomere length via a homologous recombination based mechanism (Lafferty-Whyte et al., 2009). In other cases the cells continue to proliferate even though their telomeres are dysfunctional. One example is by creating a block in the DNA damage signalling pathway preventing activation of downstream processes which would typically lead to cell death or arrest of the cell cycle (Carneiro et al., 2010).

Many pro-senescence therapeutic options are currently in development, such as the telomerase inhibitor GRN163L (Nardella et al., 2011). It is hoped that reestablishment of normal telomere dysfunction signalling in tumours, both early stage and advanced, will be an effective method of tumour reduction in cancer patients. There is also evidence that senescent cells are cleared by the immune system and this could be promising in the regression of established tumours via pro senescence therapy (Kang et al., 2011). However, the full mechanism of action, and/or signalling cascades, of many therapies targeted at the telomere is not fully understood and a clearer understanding of the underlying biology is essential if these are to be developed further.

1.3.2. Telomere dysfunction related diseases

Telomere production, regulation and associated signalling pathways include a range of biological processes. As such, there are many possible points of failure in telomere biology that can lead to a range of disease phenotypes. Mutations in many of the key functional proteins involved in telomere and telomerase regulation have been identified and are associated with a range of genetic diseases (Armanios and Blackburn, 2012; Holohan et al., 2014). These disorders often share symptoms and mechanisms and evidence suggests that they are points along a spectrum of disease. However, new causes of these disorders are continually being discovered and progress in clinical understanding of these "telomeropathies" has also renewed interest in the understanding of telomere biology.

Most of the telomerase enzyme components have been associated with telomere maintenance diseases. The first diseases associated with defective telomere maintenance, namely Dyskeratosis Congenita (DKC), were associated with mutations in the telomerase enzyme (Heiss et al., 1998). Mutations in TERT and TERC (Armanios et al., 2005; Vulliamy et al., 2001; Yamaguchi et al., 2005), Dyskerin (Heiss et al., 1998), NOP10 (Walne et al., 2007), NHP2 (Vulliamy et al., 2008), and TCAB1 (Zhong et al., 2011) have been detected in families with telomeropathies. Identification of Shelterin mutations has been more challenging. Attempts at generating mutations in Shelterin components tend to be embryonic lethal in mice models (Beier et al., 2012; Tejera et al., 2010). However, mutations in TIN2 have been identified in humans which manifest as a severe DKC disease phenotype and have early onset (Sasa et al., 2012; Savage et al., 2008; Vulliamy et al., 2012).

There have also been mutations found in other elements of telomere regulatory pathways. The t-loop must be dissociated during the DNA replication process and this is handled by RTEL1 or other helicases. When RTEL1 is absent SLX4 nuclease excises the t-loop resulting in rapid telomere shortening (Vannier et al., 2012). RTEL1 mutations have also been associated with a clinically severe form of DKC known as Hoyeraal-Hreidarsson syndrome (HHS) (Ballew et al., 2013). Associated

with t-loop formation is the 3’ overhang. Apollo, an interstrand cross-link repair nuclease is recruited by TRF2 to the telomere and is implicated in the telomere replication process (Chen et al., 2008). Depletion of Apollo, disruption of its nuclease activity or removing its ability to interact with TRF2 leads to loss in the 3’ overhang (Wu et al., 2012). Alternate splicing of Apollo, resulting in an aberrant form of the protein, has been associated in a patient with HHS suggesting that other mutations in Apollo may be responsible for other telomeropathies (Touzot et al., 2010). Other mutations which result in defective telomere replication have also been identified. The CST complex plays a key role in resolution of stalled replication forks during telomeric DNA replication and C-strand fill-in after telomerase activity during DNA replication. Mutations in CTC1, a component of the CST complex, have been associated with Coats Plus syndrome which is another severe and rare telomeropathy (Anderson et al., 2012).

Understanding the underlying biology that leads to oncogenesis, cancer progression and telomere associated diseases is essential. While there are many therapies under development targeting telomeres and telomerase in cancer (Yaswen et al., 2015), current therapy for telomere associated diseases such as DKC relies on organ or bone marrow transplants (Isoda et al., 2013; Young, 2012). Therefore further study of telomere dysfunction in cancer models and mutations which lead to telomeropathies is essential. Through further investigation of telomere biology we should be able to characterise key regulatory and signalling pathways and enable identification of targets for further therapeutic development.

1.4. Aims and Objectives

Telomere biology and regulation of telomere dysfunction is a highly complex process involving many biological pathways and signalling processes. It is known that telomere regulation is involved in oncogenesis and a range of telomere associated diseases. The overall goal of this project is to improve our understanding of telomere biology, its importance in cancer and disease and build a picture of the overall biological landscape in the context of telomere dysfunction. This will be achieved through a novel microarray visualisation and analysis strategy focusing on network-based analysis rather than gene-based analysis. Gene-based analysis tends to produce lists of genes or single genes which may have biological significance in a particular pathway. Many databases and tools help to verify *a posteriori* whether genes known to co-operate in a biological pathway are found in a list of selected genes. Basic statistical analysis then enables determination of whether a pathway is over-represented in the list, and whether it is over or under-activated. However, I would argue that introducing information on biological pathways at this point in the analysis process sacrifices statistical power to the simplicity of the approach. For example, a small but coherent difference in the expression of multiple or all genes in a pathway should be more significant than a larger difference occurring in unrelated genes. Therefore I propose that utilisation of *a priori* pathway knowledge during gene expression analysis is a more effective and simple approach to analysis of lists of genes. By taking a network focused approach it is possible to simplify the biological interpretation of microarray data and identify biological processes for further investigation.

First a new methodology will be developed in order to better visualise and interpret data comprised of gene lists followed by validation using test data sets (Chapter 3). Next the methodology will be used on a new data set to investigate the role that specific genetic lesions have in the regulation of telomere dysfunction induced by telomere uncapping and to identify candidate pathways for further analysis (Chapter 4). Then a candidate pathway (DNA Mismatch Repair) will be further characterised and specific mechanisms of action investigated (Chapter 5). Finally the methodology will then be used to

characterise alternative methods of telomere dysfunction and the synergistic effects of telomerase inhibition and hypoglycaemia, an environmental telomere dysfunction trigger identified by our group (Chapter 6).

Chapter 2: Materials and Methods

2.1. Materials

2.1.1. Viruses

Adenovirus Ad-hTRmut was supplied by Qbiogene and generated using the AdEasy system (He et al., 1998). Ad-hTRmut is based on AU5-hTer (template: 5' UAUAUAUAUAA) described by Kim et al. where the WT template (5' CUAACCCUAAC) was modified by site directed mutagenesis (Kim et al., 2001). All vectors were amplified on 293 cells and purified and quantified using BD Bioscience AdenoX virus purification and rapid titre kits according to the manufacturer's instructions.

2.1.2. Plasmids

Reporter pGL3-hTERT contains the firefly luciferase gene from pGL3 (Promega) driven by the hTERT promoter.

Reporter pLightSwitch-cMyc was supplied by SwitchGear genomics (S719321) and contains the firefly luciferase gene driven by the cMyc promoter.

2.1.3. Antibodies

Table 2.1. List of Antibodies.

Antibody	Supplier	Cat. Number
Anti-TRF2 mouse monoclonal IgG	Abcam	ab13579
Anti-CDK2 rabbit monoclonal IgG	Cell Signaling Technology	2546
Anti-MSH2 rabbit polyclonal IgG	Abcam	ab16833
Anti-MSH6 rabbit monoclonal IgG	Abcam	ab92471
Anti-MLH1 rabbit monoclonal IgG	Abcam	ab92312
Anti-ERK rabbit polyclonal IgG	Santa Cruz Biotechnology	sc-93
Anti-rabbit IgG, HRP-lined	Cell Signaling Technology	7074
Anti-mouse IgG, HRP-lined	Cell Signaling Technology	7076

2.1.4. Oligonucleotides

2.1.4.1. qPCR Primers

Table 2.2. List of qPCR Primers

Primer	Sequence	Supplier
Telomere Forward	5'-CGGTTTGTTTGGGTTT GGGTTTGGGTTTGGGTTTGGGTT	Sigma
Telomere Reverse	5'-GGCTTGCCTTACCCTTACCCTTACCCTTACCCT TACCCT	Sigma
hTERT Forward	5'-CTGCTGCGCACGTGGGAAGC	Integrated DNA Technologies
hTERT Reverse	5'-GGACACCTGGCGGAAGGAG	Integrated DNA Technologies
hTR Forward	5'-CTAACCCTAACTGAGAAGGGCGTA	Sigma
hTR Reverse	5'-GGCGAACGGGCCAGCAGCTGACATT	Sigma
p21 Forward	5'-GCTTCATGCCAGCTACTTCC	Sigma
p21 Reverse	5'-AGGTGAGGGGACTCCAAAGT	Sigma
p300 Forward	5'-ACAAATACTGCTCCAAGCTC	Sigma
p300 Reverse	5'-TAATAAGGGCATCACGCGG	Sigma
GAPDH Forward	5'-ACCACAGTCCATGCCATCAC	Sigma
GAPDH Reverse	5'-TCCACCACCCTGTTGCTGTA	Sigma

2.1.4.2. siRNA

siRNA for target genes MLH1, MSH2 or Non-targeting were all SMARTpool: ON-TARGETplus supplied by Dharmacon.

2.2. Methods

2.2.1. Tissue Culture

Human cell lines HCT116 WT, HCT116 $p21^{-/-}$ and HCT116 $p53^{-/-}$ (colon carcinoma) were maintained in RPMI1640 supplemented with 10% fetal bovine serum (FBS), 1x L-Glutamine and incubated in 5% CO2. $p21^{-/-}$ and $p53^{-/-}$ are both derived from HCT116 WT and are homozygous knockouts for p21 and p53 respectively. Glucose restriction experiments were carried out by Dr Alan Bilsland using glucose free RPMI1640 supplemented with 10% FBS.

2.2.2. Long term culture with telomerase inhibitor GRN163L

Cells were seeded into T75 flasks at 5x10^4 cells/flask on day 0. GRN163L and controls were added 3h post seeding to a final concentration of 5μM. Cultures were retreated on day 4. On day 7 the cultures were trypsinised, counted and reseeded at 5x10^4 cells/flask. 3h post seeding the cultures were treated with GRN163L and controls to a final concentration of 5μM. Cell counting was carried out manually using a haemocytometer. Treatment and counting schedule was then maintained at the same frequency over the time course. Three independent cultures were maintained for each cell line.

2.2.3. xCELLigence: Real-time growth analysis

System was blanked by adding 50 ul media to all wells of a 96-well E plate (Roche) prior to seeding. Cells were then seeded at 1000 cells per well. The viral titration of 5000ifu/cell was added 24 hours later. The system was set to record cell index every 6 hours for 5 days post infection. Three biological replicates were performed for each cell line with 3 technical replicates in each biological replicate.

2.2.4. Senescence Associated β Gal

5x10^5 cells were seeded 24 hours into T75 flasks prior to treatments. Cells were exposed to Ad-hTRmut at the IC50 determined for each cell line (Figure 4.1C) for 2 or 5 days prior to reseeding in 6 well plates at 5x10^4 cells per well followed by staining for SA β Gal. For long term culture with GRN163L, cells were seeded into 6 well plates at time of reseeding during the cumulative population doubling experiment. The SA β Gal method described by Dimri et al. was followed (Dimri et al., 1995). Three biological replicates were performed for each cell line with 3 technical replicates in each biological replicate.

2.2.5. MTT viability assay

Cells were seeded in triplicate 96-well plates 24 hours prior to treatments at 1000 cells per well. Cells were exposed to a titration of virus for 5 days prior to MTT assay. MTT reduction assays were performed using Softmax Pro 4.6 software (Molecular Devices Ltd.). For the siRNA experiments cells were transfected 24 hours post seeding with 30nM ON-TARGETplus SMARTpool siRNA (Dharmacon) using Lipofectamine2000 (Life Technologies) at a ratio of 1µl LF2000 per 3pmol siRNA. 24 hours post transfection cells were exposed to Ad-hTRmut at the IC50 determined for each cell line (Figure 4.1C) for 5 days prior to MTT assay. Three biological replicates were performed for each cell line with 3 technical replicates in each biological replicate.

2.2.6. ChIP and Telomere qPCR

Cells were seeded into T75 flasks at 5x10^5 cells per flask and then treated 24h later with Ad-hTRmut at the IC50 determined for each cell line (Figure 4.1C). After 2 days treated or untreated cells were crosslinked in 1% formaldehyde at room temperature with mild agitation. After 15 min, 0.125 M glycine was added and the cultures were agitated for another 5 min. Adherent cells were scraped off of the culture dishes with cold PBS and pelleted at 1500 rpm for 4 minutes. Cells were washed in cold PBS and re-suspended in lysis buffer 1% SDS, 10 mM EDTA, 50 mM Tris (pH 8.0). The cells were lysed on ice and then sonicated to shear the chromosomal DNA. Lysates were pre-cleared with salmon

sperm/Agarose A protein slurry (Millipore) for thirty min and the TRF2 antibody was added to the lysate plus IP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris [pH 8.0], 167 mM NaCl), and the tubes were rotated overnight at 4°C. Salmon sperm/Agarose A protein slurry was then added to the tubes for two hours, centrifuged and then liquid was aspirated. The beads were washed with successive solutions of low salt wash buffer (0.1% SDS, 1% Triton-X100, 2 mM EDTA, 20 mM Tris [pH 8.0] and 150 mM NaCl), high salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris [pH 8.0], 500 mM NaCl), LiCl2 wash buffer (250 mM LiCl2, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA and 10 mM Tris [pH 8.0]) and TE. Protein complexes were eluted off of the beads with a 1% SDS/TE solution at 65° and de-crosslinked at 65°C. After de-crosslinking, 2 µl of RNAse (0.5 mg/mL) was added to each sample and incubated for 1 hour at 37°C. Phenol:Chloroform:Isoamyl Alcohol (Sigma) was added to each sample and inverted for 2 minutes. Samples were then spun down and top layer transferred to a fresh tube. DNA was then precipitated by adding 100% ethanol and pelleted by centrifugation for 10 minutes at 13,000 rpm. Pellets were then washed in 70% ethanol. The resulting DNA pellets were re-suspended in deionised water. Telomeric DNA was then quantified by qPCR using the method described by Panero et al. using Telomere Forward and Telomere Reverse primers (Panero et al., 2015).

2.2.7. Western Blotting

Cells were seeded in 6 well plates at 5x10^5 cells per well 24 hours prior to treatments. Protein was extracted by lysing the cells in CHAPS lysis buffer on ice for 20 minutes followed by centrifugation and transfer of supernatant to a fresh tube (Millipore). Protein concentrations were estimated at OD595 using the BioRad protein assay (BioRad Laboratories Ltd, Hemel Hempstead, UK). 15µg protein per lane was then run on a 4-12% Bis-Tris Gel (Life Technologies) at 200V for 45 minutes. Proteins were transferred to a nitrocellulose membrane by electroblotting at 20V for 1 hour. Membranes were incubated in blocking solution (5% Marvel/PBS-0.2% Tween) overnight at 4ºC. Membranes were then incubated with primary antibody (in blocking solution) for 2 hours at room temperature, washed 3 times for 10 minutes in PBS-T, incubated with

appropriate secondary antibody (in blocking solution) for 1 hour at room temperature and then washed 3 times in PBS-T. Detection was carried out by applying ECL (GE Healthcare) to the membrane and then imaged with a G:Box system (Syngene). Three biological replicates were performed for each cell line with 1 technical replicates in each biological replicate. Representative blots were chosen for presentation in the appropriate results sections.

2.2.8. Microarray analysis and gene list generation

2.2.8.1. GeneSpring array data handling and GeneGo process network analysis

Cells were seeded in 6 well plates 24 hours prior to treatments. RNA was isolated with the NucleoSpin II Total RNA Isolation kit (Machery-Nagel) according to manufacturer's instructions and RNA concentration quantified by NanoDrop. RNA samples were labelled using the One-Color Microarray-Based Gene Expression Analysis Low Input Quick Amp Labelling kit and hybridised to Agilent Whole Human Genome (4x44K) arrays according to the manufacturer's instructions. All experiments were performed in triplicate (biological replicates) with a single array per biological replicate. Prior to further analysis the quality control data for each array was assessed as per the manufacturer's instructions. Only arrays which passed all quality control metrics were taken forward. Where arrays failed, additional biological replicates were performed as required. Data was extracted using Agilent Feature Extraction software version 8.1 (Agilent Technologies) and imported into GeneSpring GX 7.3.1 (Agilent Technologies) for normalisation and statistical analysis. Intra-array normalisation was carried out using the 75th percentile for each microarray. Prior to further analysis the quality of the biological replicates was assessed by principal component analysis. Of the experiments performed none had to be rejected and repeated. Significant differences in expression between control and treated cells were determined using a paired t-test and Benjamini and Hochberg false discovery rate multiple testing correction of 5%. IDs with a $p<0.005$ were selected for further analysis. The lists produced were all within the range of 500-1000 total IDs. The filtered ID lists were analysed by the MetaCore (Thomson Reuters) process networks algorithm. The process networks algorithm looks for IDs in the provided list and

matches them against an ontology of biological process networks within MetaCore, returning the most significant networks. The list of process networks were then exported to Tableau 8 (Tableau Software) for further visualisation and analysis. This process is detailed in Chapter 3.

2.2.8.2. MetaCore network building

p21, MSH2 and MSH6 were entered as a manual list in the MetaCore "build network from list" section. The Auto expand network building algorithm was then run under default settings (50 network nodes and "use canonical pathways" selected).

2.2.8.3. Validation of microarray data

Initial validation of microarray data was achieved through assessment of the individual array quality control metrics described in the manufacturer's instructions. Arrays which did not meet these quality control metrics would be rejected. However, in the experiments performed within this thesis all arrays passed this quality control step. This was followed by assessment of the sets of biological replicates in GeneSpring through principal component analysis. At this step outliers would be rejected, however no sets of biological replicates had outliers significant enough to be rejected within the experiments described in this thesis. During in silico analysis, enriched networks were determined through process network ontology analysis (PNO) which is based on a curated database of networks within GeneGo (Chapter 3). In order to further validate the networks identified a network building approach was also applied to each data set. This approach used the default stepwise network building algorithm in GeneGo which looks for direct interactions between genes in the list provided. This was used to confirm that the networks identified through the ontology analysis were in concordance with those identified through a secondary network building approach. Once specific networks were identified for further experimental investigation validation was performed. Specifically, a number of key mRNA expression changes were verified by qPCR. Further to this, the protein levels were also assessed via Western Blotting. Methodology used to perform this validation is described in this methods chapter.

2.2.9. TeloTAGGG Telomere Length Assay

Cell pellets were obtained during the long term culture at cell reseeding cycles. DNA was then extracted using the QIAamp DNA Mini Kit (Qiagen) as in the manufacturer's instructions. Sharon Burns carried out the TeloTAGGG assay for telomere length as in the manufacturer's instructions. 1 µg genomic DNA from cell pellets was digested with HinfI/RsaI. Digestion products were separated by gel electrophoresis alongside DIG-labelled molecular weight markers and blotted onto positively charged nylon membrane (Roche Diagnostics). Membranes were UV cross-linked, baked at 120°C and washed in 2xSSC solution. Hybridisation of the DIG-labelled telomeric probe was performed using buffers and probe provided. Finally, membranes were washed, probed with alkaline phosphatase conjugated anti-DIG and exposed to the CDP-star substrate. All Blots were imaged using the G:Box system (Syngene).

2.2.10. qPCR expression analysis

Cells were seeded into T75 flasks at 5x10^5 cells per flask and then treated with virus 24 hours later. Samples for RNA extraction were taken 2 and 5 days post infection and RNA was isolated with the NucleoSpin II Total RNA Isolation kit (Machery-Nagel) according to manufacturer's instructions. cDNA was produced using 1ug input RNA with the GeneAmp RNA PCR Core Kit (Applied Biosystems). Q-PCR was performed using DNA Engine Opticon 2 equipment and software (BioRad). SYBR green was used as fluorophore (ThermoFisher). Optical read temperatures were optimised to exclude primer dimers.

2.2.11. Luciferase promoter assay

Cells were seeded into 96 well plates at 1000 cells per well then transfected 24 hours later with 250ng pGL3-hTERT or pLightSwitch-cMyc reporter plasmid with superfect at a ratio of 2.5:1 superfect:DNA. Virus was added 3 hours post transfection at the IC50 determined for each cell line (Figure 4.1C). The cells were incubated for 48 hours and then luciferase activity was determined using luciferase assay reagents according to the manufacturer's instructions

(Promega). Three biological replicates were performed for each cell line with 3 technical replicates in each biological replicate.

2.2.12. TRAP assay for telomerase activity

The TRAPeze XL kit was used for TRAP assay according to the manufacturer's instructions (Millipore). Cell pellets were lysed in CHAPS lysis buffer and protein concentrations estimated by Bio-Rad assay (BioRad). 0.5 µg protein was mixed with TRAPeze reaction mix containing TS primer, fluorescein labelled RP primer, control template and sulforhodamine labelled control K2 primer. Each assay included no-telomerase, no-Taq, and heat- treated controls. Extension products were generated at 30°C followed by Q-PCR detection using Chromo4 equipment and software (BioRad). Total product generated was measured against TR8 standards and normalised to the ROX internal control. Three biological replicates were performed for each cell line with 1 technical replicates in each biological replicate.

2.2.13. Statistical Analysis

For statistical analysis, unpaired, two-tailed Student's t-tests were used and carried out in Microsoft Excel. P values of <0.05 were considered significant.

Chapter 3: A novel methodology for visualisation and analysis of data sets comprised of gene lists

3.1. Introduction

Traditionally, research approaches in biology focus on one or a few genes at a time. However, due to advances in various technologies such as expression microarrays and bioinformatics analysis techniques it is possible to simultaneously measure genome-wide changes under various biological conditions and timeframes. This results in the generation of large gene lists and vast total data sets of which interpretation and visualisation is still a challenging task. Fortunately, over the last few decades there have also been advances in the bioinformatics methods and public databases used to interrogate these data sets making it possible to draw conclusions and identify relevant biology. Many enrichment tools have been developed in order to assist in this process in order to address the problem is analysing such large data sets. For example, in 2008 Huang et al. carried out a review of 68 of these enrichment tools, highlighting how quickly the field has grown and the diversity of tools available in assisting investigators in analysing their data sets (Huang da et al., 2009).

However, although enrichment analysis approaches and tools have developed significantly over the years, visualization and interpretation of these large data sets is still a challenging task. Although tools are available to visualise individual data sets or small groups of data it is still challenging to efficiently compare a wider range of data sets and query them in a sensible manner. A wide range of software tools have also been developed to assist in visualization. However, while some are fairly mature, such as molecular graphics (O'Donoghue et al., 2010b) others are less so, such as visualisation of genomes (Nielsen et al., 2010). Also, as software has advanced so has the hardware, to the point at which many bioinformatics analysis techniques can be carried out with a standard personal computer making it an accessible and useful option for many investigators. Integration of a variety of visualization tools is a desirable goal for investigators working with large data sets. Modern visualization tools, such as Jalview, are often integrated with remote databases that allow for visualizations that

integrate data from multiple sources (Waterhouse et al., 2009). In addition, many of these tools are designed in order to integrate directly with other visualization and analysis tools. For example, it is possible to visualise multiple sequence alignments whilst comparing the corresponding three dimensional structures (O'Donoghue et al., 2010b; Procter et al., 2010) or of a network with corresponding heat maps, profile plots or phylogenetic trees and dendrograms (Gehlenborg et al., 2010). However, many of these integrated visualizations are limited in their ability to compare multiple data sets efficiently and easily.

As our lab generated array data, it became essential to manage these data in a way which allowed for multiple data sets to be easily retrieved and compared so that meaningful analysis could be made. Traditional microarray expression analysis often focuses on identifying small subsets of genes which may play a role in a specific biological process (Huang da et al., 2009). However, this approach is limited in that the genes identified may not be directly connected in any kind of pathway. We desired a method which allowed for identification of specific pathways within our data sets, rather than simply lists of genes. Previous work within our group has been to utilise MetaCore, a database and software package for analysis of gene lists, in order to analyse our data sets (Bilsland et al., 2009; Ekins et al., 2006). In order to develop our analysis methods further my approach has been to use an existing functional ontology analysis available within MetaCore, called the process network ontology (or PNO) followed by a novel visualisation and analysis method. This has allowed us to extract meaningful conclusions primarily from microarray expression data but also from other gene list sources. The process network ontology is made up of 169 networks covering 23 top-level biological processes shown in Table 3.1, such as Cell Cycle and DNA damage, which are curated and updated based on current literature. The goal was to develop a workflow and visualisation process that would enable accurate and time efficient comparisons of experimental array data within our database as a whole. It was also desired that the methodology could be flexible in order to be adapted to other data sets involving large lists of genes.

3.2. Results

3.2.1. Novel expression analysis methodology enables identification of specific pathways within data sets

In order to maintain consistency between data sets so that comparative analysis could be made, a workflow was developed for data handling and analysis shown in Figure 3.1. For internal data sets, expression data was generated using the Agilent 44k whole genome microarray platform in triplicate. The data generated was then taken into GeneSpring GX12 for data normalisation, quality control and statistical analysis. This produced filtered gene lists based on a significant p value (such as $p<0.05$) and optionally, fold change, for further analysis. External data could be integrated at this point; in the case of raw data, or directly into the next step of analysis. The novel analysis workflow could then be performed from this point on. First the gene list is processed in MetaCore where functional enrichment analysis is performed on the list comparing it against the process network ontology (or PNO). This analysis returns a list of 169 networks, covering 23 top-level biological processes shown in Table 3.1, ordered by p value. This list is then entered into a database along with all other PNO data generated. After this point data is visualised in Tableau, a software package that allows real time visualisation of database information in a variety of graphical formats. During visualisation, differentially enriched biological processes and networks are identified which can be further investigated. The data can also be further processed into a clustering diagram to directly quantify the similarity between different data sets.

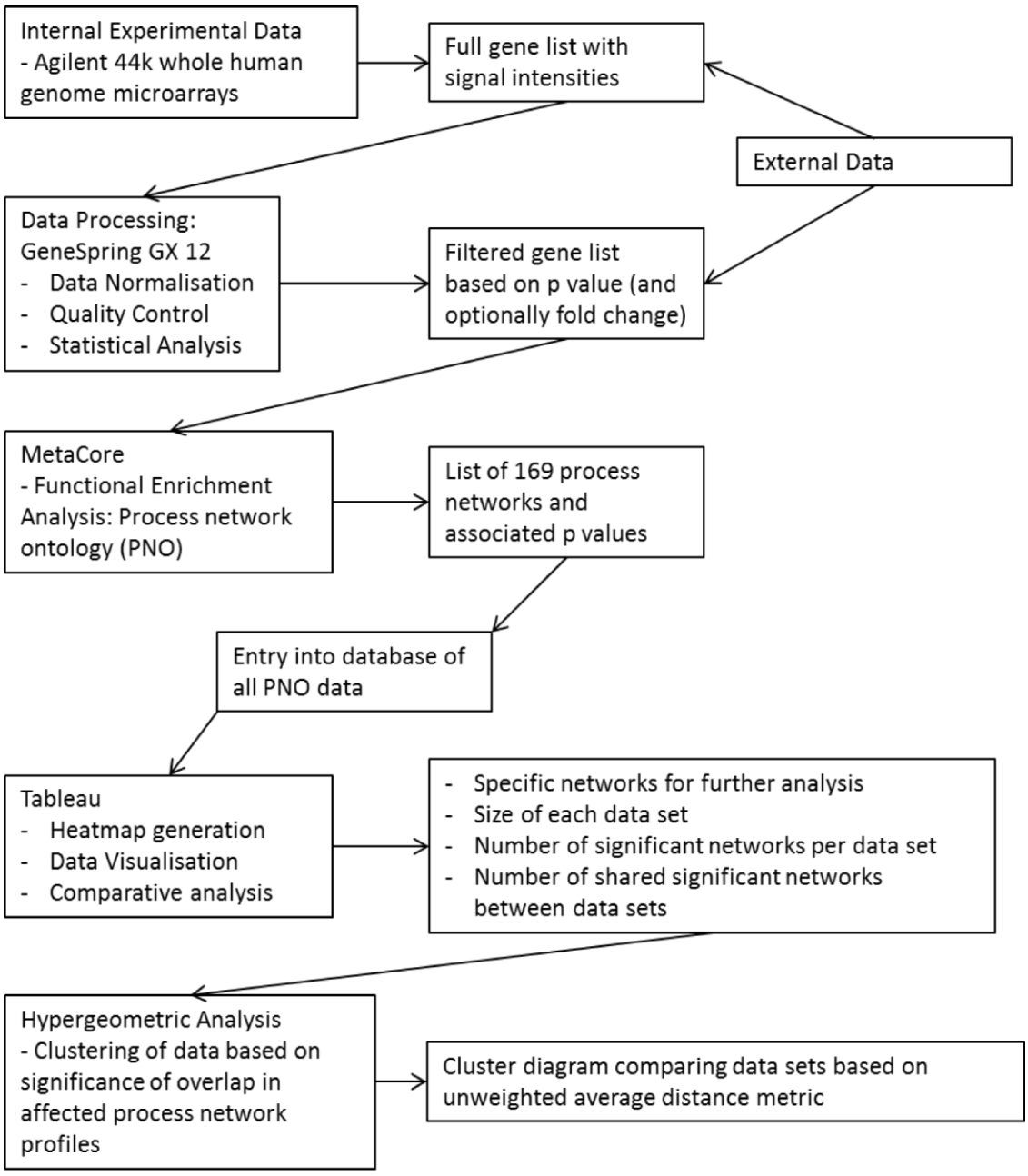


Figure 3.1. Novel workflow for gene list analysis focusing on network and pathway identification.

Processes	Number of process networks
Apoptosis and survival	10
Autophagy	1
Blood coagulation	1
Cell adhesion	12
Cell cycle	11
Chemotaxis	1
Cytoskeleton remodelling	6
Development	20
DNA damage	5
Immune response	10
Inflammation	21
Muscle contraction	2
Neurophysiological process	11
Proliferation	3
Protein folding	4
Proteolysis	4
Regulation of metabolism	1
Reproduction	10
Response to hypoxia and oxidative stress	2
Signal transduction	17
Transcription	4
Translation	5
Transport	8

Table 3.1 MetaCore process network ontology.

Each process consists of a set of curated signalling networks which are based on current literature and updated quarterly.

Shown in Figure 3.2 is a more detailed diagram of the visualisation process where three HCT116 isogenic cell lines (Cell line A: WT, Cell line B: $21^{-/-}$ and Cell line C: $p53^{-/-}$) are being compared after induction of telomere dysfunction by comparing hTRmut treated samples with mock infected samples. Figure 3.2A demonstrates the top-level profile view generated in Tableau. Red rectangles are networks where $p<0.05$ indicating significance while light green to dark green are the range from $p>0.05$ to $p = 1$. At this stage it is possible to include additional data sets in the comparison as desired. For example, if there was a time course carried out it would be possible to compare time points for a single cell line, compare a single time point for multiple cell lines (as shown here) or any combination desired within the interface. This makes is quick and efficient to bring together multiple data sets for interpretation and analysis. Once a comparison is chosen it is then possible to "zoom" further in to the data set in multiple steps. The first step is shown in Figure 3.2B. In this example the Cell cycle top-level process was identified as having a number of differentially enriched processes between the different cell lines. Further assessment, with the specific network names shown, revealed that Cell Line A demonstrated enrichment for almost all of the networks available for cell cycle, Cell Line B had fewer enriched networks but did share two with Cell Line A, and Cell Line C only showed enrichment in a single network which was not enriched in the other two cell lines. The next step of increasing resolution on the data is shown in Figure 3.2C which shows the cell cycle core network. At this point it is possible to assess specific expression changes within the network and identify possible signalling pathways for further investigation. Finally, shown in Figure 3.2D, the approach of clustering responses based on significance of overlap in affected process network profiles was adopted and implemented by Dr Alan Bilsland. The number of significant process networks affected in each column in Figure 3.2A and the number shared in each pairwise comparison were used to generate cumulative hypergeometric probabilities for each pair, which was used as an unweighted average distance metric. This allows the level of similarity between profiles to be quantified and visualised in a clustering diagram.

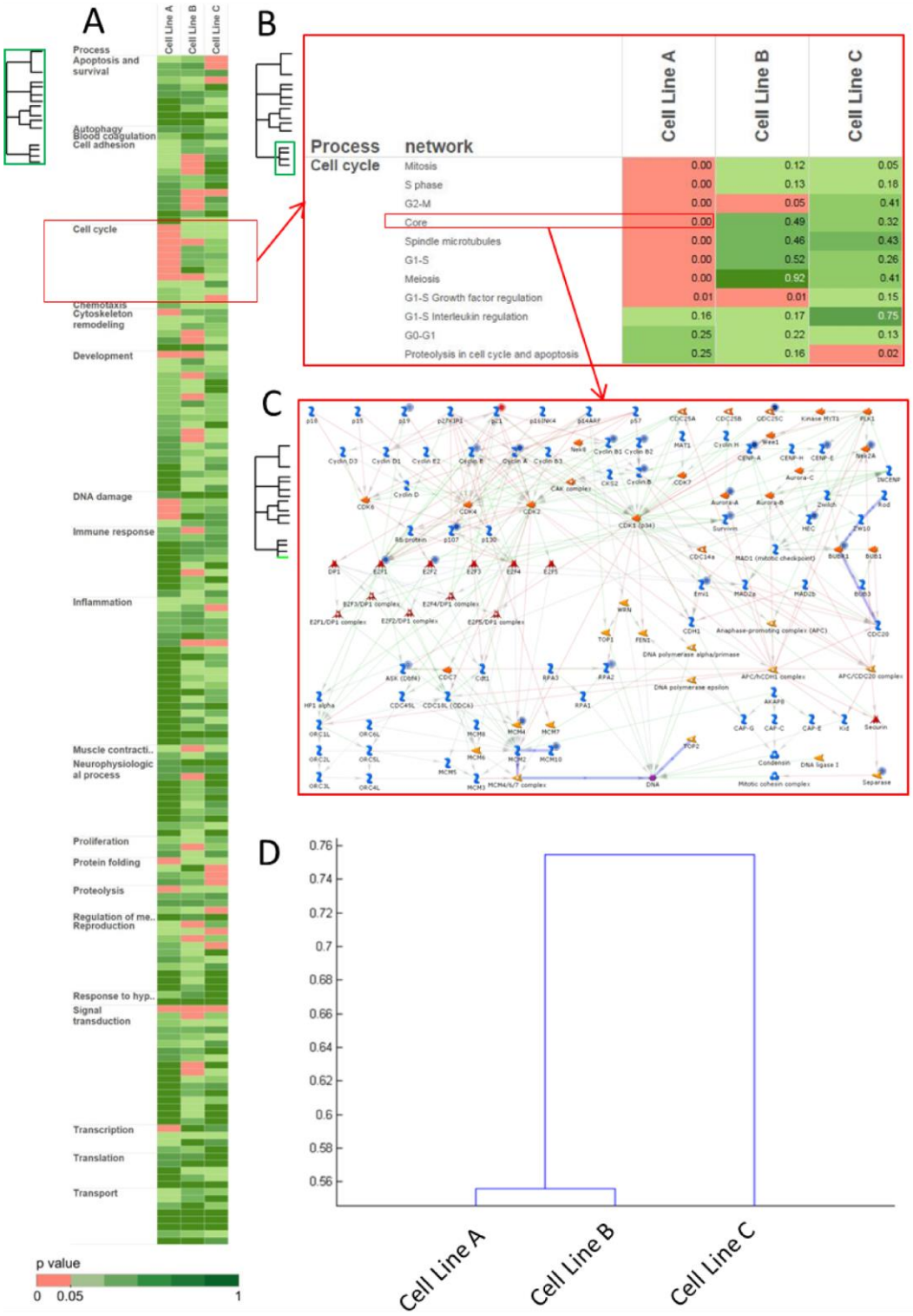


Figure 3.2. Visualisation and analysis of MetaCore process network ontology profiles.

(A) Process network profile of all 169 networks comparing three isogenic HCT116 cell lines after induction of telomere dysfunction in a heatmap format. Significant process networks ($p<0.05$) are indicated as red in the heat map. (B) Subset view showing all the process networks in the top level biological process cell cycle. (C) Cell cycle core as an example of an individual process network as output of the methodology. (D) Clustering of process network profiles on cumulative hypergeometric probability of pairwise overlap.

3.2.2. Characterisation of mechanism of action of a novel senescence inducing drug

This work was carried out in collaboration with Dr Alan Bilsland, University of Glasgow and published in Neoplasia (Bilsland et al., 2015).

One of the major aims of the methodology was to allow for convenient and meaningful comparisons between different data sets within the database. In order to test my workflow model I worked with Dr Alan Bilsland to assist him in screening of senescence inducing drugs. Dr Alan Bilsland carried out a novel screening approach using neural network based learning algorithms to screen a large database of drugs for potential senescence inducing targets. Dr Bilsland reasoned that identification of enriched libraries would be beneficial before initiating a screening campaign and that virtual screening might identify such an enriched set. Bilsland et al built profiles of differentially affected biological process networks from expression data obtained under induced telomere dysfunction conditions in colorectal cancer cells and matched them to a panel of 17 protein targets with confirmatory screening data in PubChem. They then trained a neural network using 3517 compounds, identified as active or inactive, against these targets. The resulting classification model was used to screen a virtual library of ~2M lead-like compounds. One hundred and forty-seven virtual hits were acquired for validation in growth inhibition and senescence-associated ß-galactosidase assays. Among the found hits, a benzimidazolone compound, CB-20903630, had low micromolar IC50 for growth inhibition of HCT116 cells and selectively induced senescence-associated ß-galactosidase activity in the entire treated cell population without cytotoxicity or apoptosis induction. Growth suppression was mediated by G1 blockade involving increased p21 expression and suppressed cyclin B1, CDK1, and CDC25C. In addition, the compound inhibited growth of multicellular spheroids and caused severe retardation of population kinetics in long-term treatments.

During the screen a novel compound was identified, CB-2090363, which induced SA-ßGal staining in HCT116 cells and reduced cell viability over a short time course indicating senescence induction. CB-20903630 contains a kinase hinge-

binding motif (Liao, 2007), suggesting that the compound may target a cell-cycle related kinase. Expression profiles were previously generated of IMR90 fibroblasts treated with 13 well-characterised kinase inhibitors shown in Table 3.2 which induce heterochromatin foci. Most of this inhibitor panel mediate effects on the cell cycle in cancer cell lines and induce both apoptosis and senescence responses. Dr Alan Bilsland provided me with the expression profile dataset which I then compared using the PNO analysis method in order to identify affected biological pathways which may then inform on the mechanism of action of the compound (Figure 3.3A). These expression profiles were also compared against a "target profile" which was produced by confirmatory PubChem compound screens identified against a panel of 17 protein targets suitable to build a "senescence-like" profile. The heatmap demonstrates that a number of the compounds share similarity with the target profile with specifically enriched networks in Cell Cycle and DNA Damage response which would be expected under senescence. To then identify any similarity of the compound CB-2093630 to any of these kinases the compound was included in the previous analysis and a clustering diagram was produced using the hypergeometric analysis method. The resulting diagram is shown in Figure 3.3B and demonstrates that the profile for CB-2093630 clustered away from the kinases in the panel. This could indicate a mechanism of action for this compound which is not strongly dependent on the primary targets of the other inhibitors. In the group of kinase inhibitors with which CB-20903630 was being compared, cell cycle pathways, cell adhesion, and developmental and cytoskeletal processes were affected. Two AKT inhibitors (AKTV and AKTVIII) were also present in the analysis, alongside two PI3K inhibitors (PI103 and GDC0941). AKTV/GDC0941 clustered together and adhesion, inflammation, development, and proteolysis processes are strongly represented in this group. MAPK inhibitor MK2A clustered with Src-family inhibitor SU6656 and with AuroraII, primarily affecting DNA damage, cell cycle, and apoptosis processes. The CB-20903630 process network profile clustered away from all others, possibly indicating that the compound mechanism affects different pathways in normal versus cancer cells. Inflammatory processes were also highly represented in CB-20903630 treated cells. However, CB-20903630 treated cells also scored highly in a range of development and proteolysis processes shared by the

PI3K/AKT agents, making the observed profile more similar to these. Thus, different pathways may be affected by the compound in normal versus cancer cells and this analysis methodology allowed for identification of a range of potential pathways for further research. Overall the pathways identified were in concordance with pathways one would expect to find enriched in response to senescence inducing stimuli.

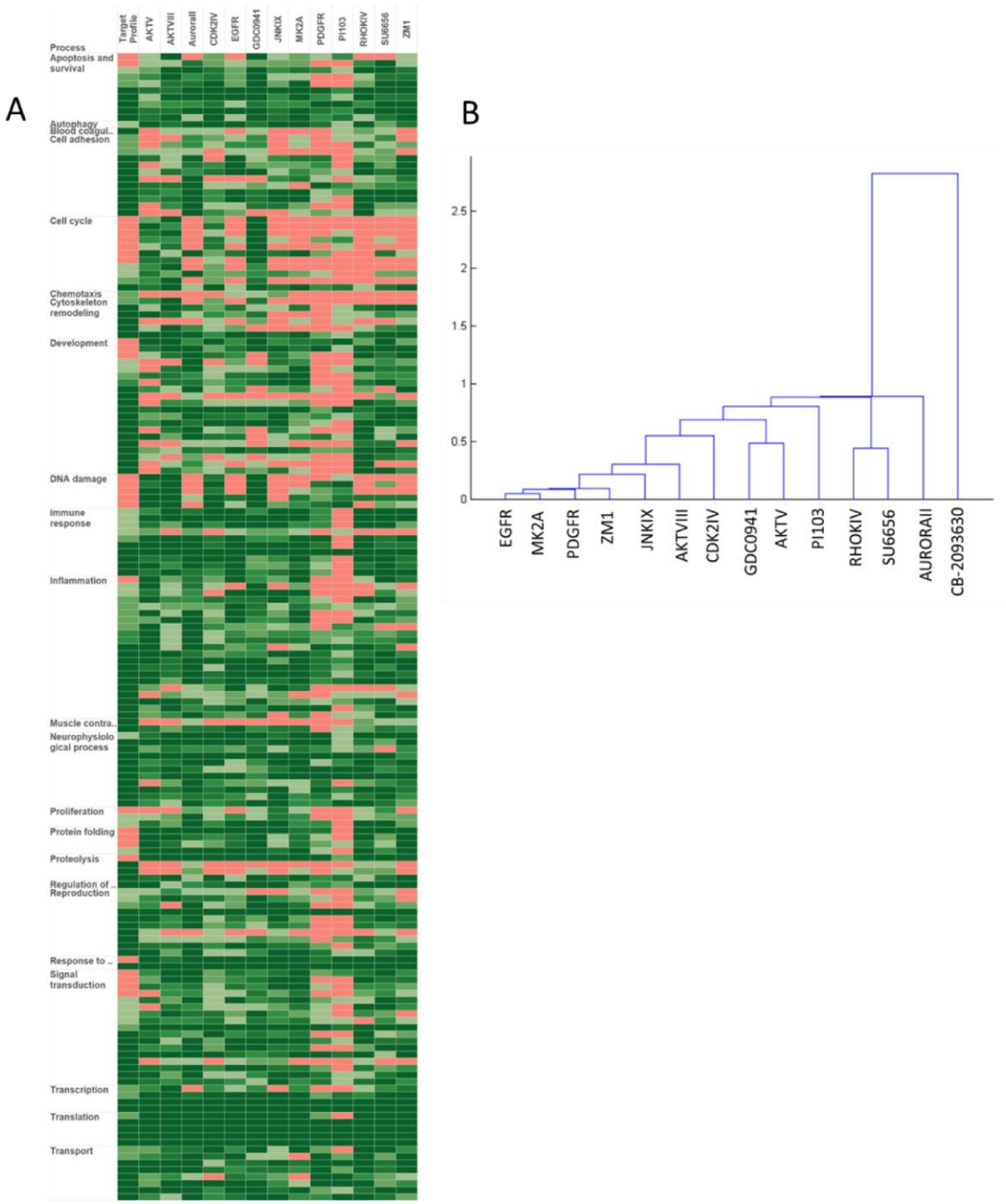


Figure 3.3. Comparison of novel senescence inducer CB-2090363 against a range of kinase inhibitors.

(A) MetaCore process network ontology analysis of selected kinase inhibitor compounds listed in Table 3.2. 3 independent RNA samples from DMSO versus compound treated cells were profiled on Agilent whole genome expression arrays. (B) Hypergeometric analysis of Figure 3.3A quantifying the similarity between the process network profiles including CB-209003630 treated cells.

Compound	Class	Targets	Millipore #
EGFR	4,6-dianilinopyrimidine	EGFR	324674
MK2A	p-amidophenol	MAPKAPK2a	475863
PDGFR	bis (1H-2-indolyl)-1-methanone	PDGFR/Flt-3	521230
ZM1	Quinazoline	Aurora A/B	189410
JNKIX	Thienylnaphthamide	JNK2/3	420136
AKTVIII	Quinoxaline	AKT1/2	124018
PI103	Pyridinylfuranopyrimidine	DNA-PK/PI3K/mTOR	528100
CDK2IV	Purine	CDK2 < CDK1/4/5/7	238804
GDC0941	Thienopyrimidine	PI3K	509226
AKTV	Tricyclic nucleoside	AKT1/2/3	124038
AuroraII	Anilinoquinazoline	Aurora A/B	189404
RhoKIV	Glycyl-isoquinolinesulfonamide	ROCK	555554
SU6656	Indolinone	Src/Fyn/Yes/Lyn	572635

Table 3.2. Signal transduction inhibitors used in process network profile clustering.

3.2.3. Identification of pathways involved in the immortalisation process in T cells

This work was carried out in collaboration with Dr Sofie Degerman, Umeå University and published in Neoplasia (Degerman et al., 2014).

While the primary intent of the methodology was for analysis of expression data it was realised that it could be applied to most data sets involving lists of genes. In order to test this I collaborated with Dr Sofie Degerman and her group to assist in analysis of an alternative data set involving lists of differentially methylated genes. In this example methylation changes during the immortalisation process were analysed by comparing methylation state of cell lines over a time course during which growth crisis was bypassed. Over the time course, lists of genes which were differentially methylated were generated which could be compared and analysed by PNO analysis. The differential methylation ($\delta\beta$) was defined as the largest change between post- and pre-crisis samples in the cell cultures. CpG-sites with a $\delta\beta$ greater than 0.4 or less than -0.4, were classified as differentially methylated (DM-CpG). Degerman et al investigated a model of spontaneous immortalization of T-cells to explore the role of genome-wide methylation in the immortalization process at different time points pre-crisis and post-crisis using high-resolution arrays. Degerman et all demonstrated that over time in culture there is an overall accumulation of methylation alterations, with preferential increased methylation close to transcription start sites (TSSs), islands, and shore regions. Methylation and gene expression alterations did not correlate for the majority of genes, but for the fraction that correlate, gain of methylation close to TSS was associated with decreased gene expression. Interestingly, the pattern of CpG site methylation observed in immortal T-cell cultures was similar to clinical T-cell acute lymphoblastic leukemia (T-ALL) samples classified as CpG island methylator phenotype positive (CIMP+). Degerman et all analysed data from 10 diagnostic pediatric T-ALL samples (7 CIMP+ and 3CIMP−) with two independently immortalised T-cell cultures (S3R and S4) using traditional array analysis clustering methods. Hierarchical clustering of the cell cultures and leukemias separated the CpGs into three clusters; cluster 1 in which de novo methylated CpGs in the immortal cells overlapped with methylated CpGs in CIMP+ T-ALL

diagnostic samples; CpGs in cluster 2 were de novo methylated in immortal cell cultures but less methylated in leukemia, and cluster 3 demethylated in immortal cell cultures but methylated in leukemia. Further to this they identified that there was a significant overrepresentation of Polycomb target genes (PCTGs) compared to random methylation in all clusters. However, the overrepresentation was most evident in cluster1 where 51% of the CpG sites were located in in PCTGs compared to 23%and 21% respectively, in clusters 2 and 3. Furthermore, the majority (62%) of the shared CpG sites in immortal cell cultures and CIMP+ leukemias were located in CpG islands.

Dr Degerman provided me with their list of differentially methylated genes from cluster 1 for further analysis in order to identify pathways for further investigation. Shown in Figure 3.4A is the initial profile generated by comparing PCTGs (Lee et al., 2006), two independently immortalised T cell cultures S3R and S4 over at post crisis time points and CpG island methylator phenotype negative or positive cell lines (CIMP- or CIMP+). This initial comparison highlighted a number of overlapping biological processes and shared networks between profiles. The process networks of highest relevance were cell adhesion, cytoskeletal remodelling, development, and signal transduction processes (Figure 3.4B). Interestingly, a large set of genes in the Wnt signalling pathway became methylated during immortalization of cell cultures and in diagnostic CIMP positive leukaemia (Figure 3.4C). Overall the methodology allowed for novel comparison of gene lists produced from methylation data and provided candidate pathways for further analysis which may be involved in the immortalisation process.

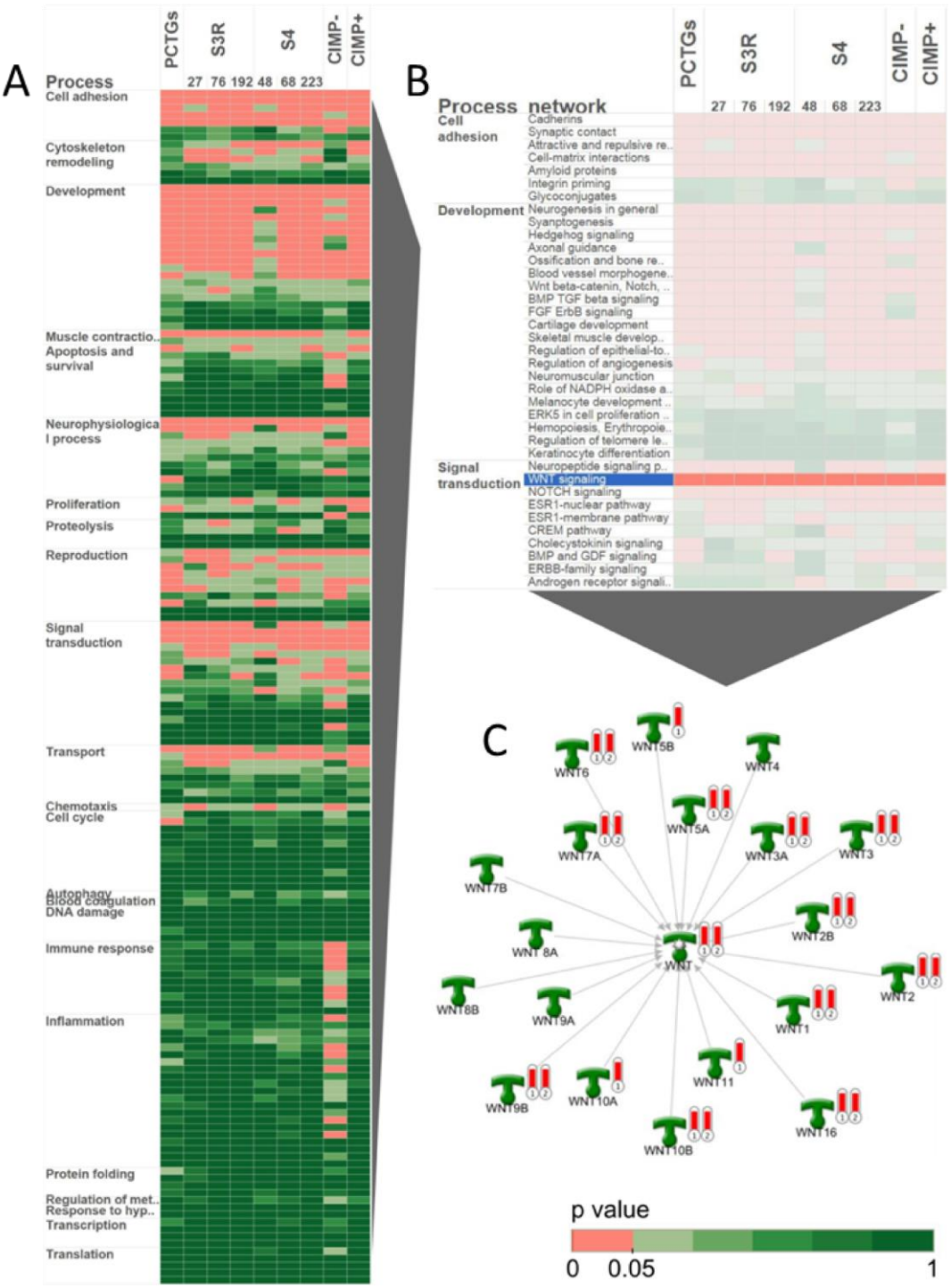


Figure 3.4. Metacore process network ontology analysis of cellular processes affected by methylation.

(A) MetaCore process network ontology analysis of differentially methylated CpG sites between post-crisis T cell cultures (S3R and S4) and CpG island methylator phenotype negative or positive cell lines (CIMP- or CIMP+). CIMP- (n=3) and CIMP+ (n=7) samples were grouped in the analysis whereas each time point are shown in for the cell cultures. Significant process ($p<0.05$) is indicated in red in the heat map. (B) The most highly enriched processes identified in the process network analysis. (C) WNT signaling genes representing a candidate process with many affected genes across all of the cell lines. The red thermometers represents 1) Polycomb target genes (PCTGs) 2) Shared CpG sites post crisis T cell cultures/CIMP+ T-ALL.

3.3. Discussion

3.3.1. Novel methodology allows for discovery of new biology

Traditional methodology typically identifies single genes or sets of genes for further investigation. It is then possible to verify *a posteriori* whether genes known to co-operate in a biological pathway are found in a list of selected genes. Basic statistical analysis then enables determination of whether a pathway is over-represented in the list, and whether it is over or under-activated. However, I would argue that introducing information on biological pathways at this point in the analysis process sacrifices statistical power to the simplicity of the approach. For example, a small but coherent difference in the expression of multiple or all genes in a pathway should be more significant than a larger difference occurring in unrelated genes. Therefore I propose that utilisation of *a priori* pathway knowledge during gene expression analysis is a more effective and simple approach to analysis of lists of genes. By using a network based ontology for analysis of our data sets it has been possible to identify specific pathways and signalling processes for further investigation in order to reveal new biology in an area of study. Another advantage of the methodology employed is that it expands on the base functionality of the MetaCore functional ontology enrichment by improving on visualisation of the data and the ability to compare multiple data sets in a clustering format. Within MetaCore, comparative analysis has to be re-run for any comparison that is being made. However, our method allows for multiple sets of data to be visualised in real time within Tableau software. This was demonstrated in two different data sets. Firstly with gene lists generated from gene expression microarray data and secondly from gene lists generated from methylation array data. In both cases the methodology allowed for comparison, visualisation and analysis of the gene lists in order to identify specific networks of interest for further study and to come to meaningful conclusions about the biology involved in each case.

In the first example, the methodology allowed a novel senescence inducing compound to be compared with a range of kinase inhibitors with which it was thought may share a mechanism of action. Using traditional methodology, Dr

Bilsland would have been able to cluster the expression profiles for each of the kinase inhibitors and produce a similar clustering diagram as shown in Figure 3.3B. However, this would have been based on clustering lists of genes rather than by clustering on networks of genes. Therefore, in order to identify mechanisms of action from such a data set further analysis would be required. By using a network focused approach this is simplified, as the data set intrinsically identifies networks of interest for further study. Indeed, clustering during the PNO analysis method revealed that while many of the kinases in the panel did show similarity the drug compound of interest appeared to cluster away from the other kinases indicating that it may induce senescence through an alternative mechanism. In addition, specific networks were identified within the clusters for further analysis. For example, CB-20903630 treated cells scored highly in inflammation, development and proteolysis processes, some of which were unique to the CB-20903630 treated cell profile and could be a direction for further study.

In the second example, the methodology was demonstrated to be flexible and applicable to gene lists generated from a different methodology. Using a traditional array analysis methodology, Degerman et al were able to identify 2 clusters of genes within their data set and demonstrate regions of overlap between the cell lines examined in their study. They were also able to demonstrate at this stage a link between the immortalisation process and polycomb target genes (PCTGs). However, at this stage specific biological processes were not identified for further study or as a potential explanation for the results obtained. PNO analysis revealed that two spontaneously immortalised T cell lines had significant overlap in enriched networks with PCTGs and CIMP+ or CIMP- cell lines and enabled identification of specific biological processes that may be involved in the immortalisation process and be specifically associated with changes in methylation status of the cells. This analysis also suggested that genes which become methylated during the immortalisation process may be specific and that the processes affected by methylation are consistent across T cell lines. These shared processes were involved in cell adhesion, cytoskeleton remodeling, development and signal transduction. Of special notice was a high number of methylation altered genes in the Wnt signaling pathways. The reason

for this is unclear but indicates selectivity in the methylation targets. In addition the analysis helped to reveal that targets of the polycomb may be important in the immortalisation process and be preferentially methylated as a result. As methylation and the immortalisation process is not a well characterised area of research I could not confirm if the pathways identified were in concordance with current theory. However, the traditional methodology identified that there was 51% of the CpGs in cluster 1 were PCTGs. Using the PNO method this remained consistent and at the network level there was a large amount of overlap between the PCTG list of networks and the list of networks generated for the S3R, S4, CIMP+ and CIMP- lists. Therefore, the methodology appears to produce results consistent with traditional methodology whilst providing specific information about biological networks which may be of value for future research.

3.3.2. Areas for further development

One weakness of the method used was that there were some bottlenecks in the workflow to steps which required manual handling of the data. The list of process network profiles generated in MetaCore had to be manually exported then parsed by an excel template to produce a formatted list which could be entered into a database before it could be queried and visualised in Tableau. The process, though relatively simple, could be rather time intensive, depending on the number of lists being processed. MetaCore now has a developer site where users can query their data directly via an application program interface (API). In theory, one could upload data to MetaCore and then query the data via the API automatically, removing the need to export manually, thus increasing the efficiency of the process while also reducing chances for user error. Also, in order to carry out the hypergeometric analysis and generate clustering diagrams the heat maps had to be compared by hand. They could then be run through a MATLAB script, produced by Dr Alan Bilsland, in order to generate the clustering diagrams. This step could also be automated in order to further streamline the process.

Another limitation is in the ability to easily query public data bases for information, particularly for sourcing external data sets. There are many public

databases available which contain sets of expression data with which automated data retrieval could be implemented. Examples for public array databases include the Gene Expression Omnibus (GEO) and ArrayExpress (Barrett and Edgar, 2006a; Parkinson et al., 2007). There are also a range of other public databases for different data sets which may be compatible with our current database and complement the gene expression data set, such as protein-protein interaction data from the IntAct database (Hermjakob et al., 2004). However, whilst databases such as GEO are accessible and can be interacted with by non-specialists thanks to web based tools and applications, the process still would require manual input into the database (Barrett and Edgar, 2006b). Similar to using API with MetaCore, many services have methods from which their databases can be queried automatically. GEO offers programmatic access to its database. Theoretically, it would be possible to develop a tool which could extract information, parse the data generated and incorporate it into our database in an automated fashion.

Chapter 4: Regulation of telomere dysfunction is influenced by genetic background

4.1. Introduction

Telomeres are regulated and affected by a wide range of biological processes, ranging from DNA damage signalling to metabolism and oxidative stress. Due to the many signalling routes engaged by the telomeres, under normal conditions it is challenging to experimentally identify key processes. Single mutations within the signalling pathways can lead to very different phenotypic outcomes. For example, yeast cells lacking taz1 (TRF1 and TRF2 human orthologue) are unable to recruit Crb2^{53BP1} and do not initiate a DNA damage response after induction of telomere dysfunction (Carneiro et al., 2010). However, by comparing the outcomes of specific mutations it is possible to further understand the complex processes that are engaged as a result of telomere dysfunction.

Ad-hTRmut is an adenovirus construct which allows for expression of a mutant variant of hTR, the RNA template component of telomerase enzyme (Kim et al., 2001; Mahalingam et al., 2011). When hTRmut is expressed in cells positive for telomerase activity it becomes incorporated into the telomerase enzyme and leads to the incorporation of mutations into the telomeric sequence. As the Shelterin proteins bind specific telomeric sequences this prevents these proteins from binding and protecting the telomeres. In turn this leads to telomere uncapping. Unprotected telomeres are exposed to endogenous DNA damage detection proteins and lead to signalling of many biological pathways, typically leading to cell cycle arrest (senescence) and/or apoptosis (Palm and de Lange, 2008). However, when specific genetic lesions are present this can lead to alterations in the signalling and resulting regulatory responses to telomere uncapping. For example, p21 and p53 are known to play key roles in DNA damage signalling and cell cycle progression in response to telomere dysfunction (Salama et al., 2014). By using cell lines without these key genes and studying these changes in signalling it is possible to gain a better understanding of how telomere dysfunction is regulated in cancer cell lines. The aim of this chapter is to characterise the outcome of telomere dysfunction in isogenic cell lines

(HCT116 WT, $p21^{-/-}$ or $p53^{-/-}$) after induction of telomere uncapping. This chapter will also identify specific pathways involved in telomere dysfunction for further analysis.

4.2. Results

4.2.1. Loss of p21 enhances sensitivity to telomere uncapping

In order to characterise the effect of specific genetic lesions on the outcome of telomere uncapping, a panel of isogenic HCT116 cell lines were chosen: WT (Wild Type), $p53^{-/-}$ and $p21^{-/-}$. The initial approach was to study the growth phenotype of these cell lines to identify any immediate differences in response to the induction of telomere dysfunction. First, the real time growth rate of the cells was measured using the xCELLigence platform (Roche) over a 144 hour period (Figure 4.1A). 5000 ifu/cell of Ad-hTRmut was added to the cells 24h post seeding and fold change in cell index from the 6h time point was measured. Variability in sensitivity to telomere uncapping by Ad-hTRmut can be seen across HCT116 $p21^{-/-}$, $p53^{-/-}$ and WT cells. WT, $p53^{-/-}$ and $p21^{-/-}$ peaked at 14, 11 and 3 fold increase in cell index respectively. However, the cell lines reached their peaks at different time points. WT and $p53^{-/-}$ both peaked at ~110 hours whereas p21-/- peaked earlier at ~90 hours. Overall, $p53^{-/-}$ and $p21^{-/-}$ cells were more sensitive to rapid telomere uncapping than WT. $p21^{-/-}$ cells demonstrated the greatest sensitivity with the lowest peak in growth over the time frame, slowest rate of growth and earliest peak in growth indicating that $p21^{-/-}$ cells are affected more quickly by telomere uncapping.

A common outcome of telomere dysfunction is cellular senescence. To characterise the senescent population of cells after induction of telomere dysfunction an SA ß Gal assay was carried out. This assay measures ß-galactosidase levels which is a senescence marker. Cells were treated with Ad-hTRmut at the IC50 shown in Figure 4.1C. As shown in Figure 4.1B, there was no change in the senescent population of cells relative to mock infected cells in any of the lines at day 2 post treatment with Ad-hTRmut. However, by day 5 the senescent population had increased 7 fold in WT and $p53^{-/-}$ and 5 fold in $p21^{-/-}$

cells. This may suggest weakened or delayed senescence signalling in $p21^{-/-}$ cells in response to telomere dysfunction triggered by Ad-hTRmut relative to the other lines.

To further investigate the growth phenotype of the cells an MTT viability assay was carried out. In the MTT assay in Figure 4.1C this range of sensitivity was also observed. All cell lines responded to Ad-hTRmut in a dose dependent manner and WT, $p53^{-/-}$ and $p21^{-/-}$had an IC50 of ~4800, ~3000 and ~420 ifu/cell respectively. In order to confirm that the effects seen could be attributed to telomere uncapping a ChIP assay was performed for TRF2, a known Shelterin component binding to telomeric DNA. Day 2 post treatment with 5000 ifu/cell Ad-hTRmut, it was found that TRF2 binding to telomeric sequence was reduced significantly in all cell lines (Figure 4.1D).

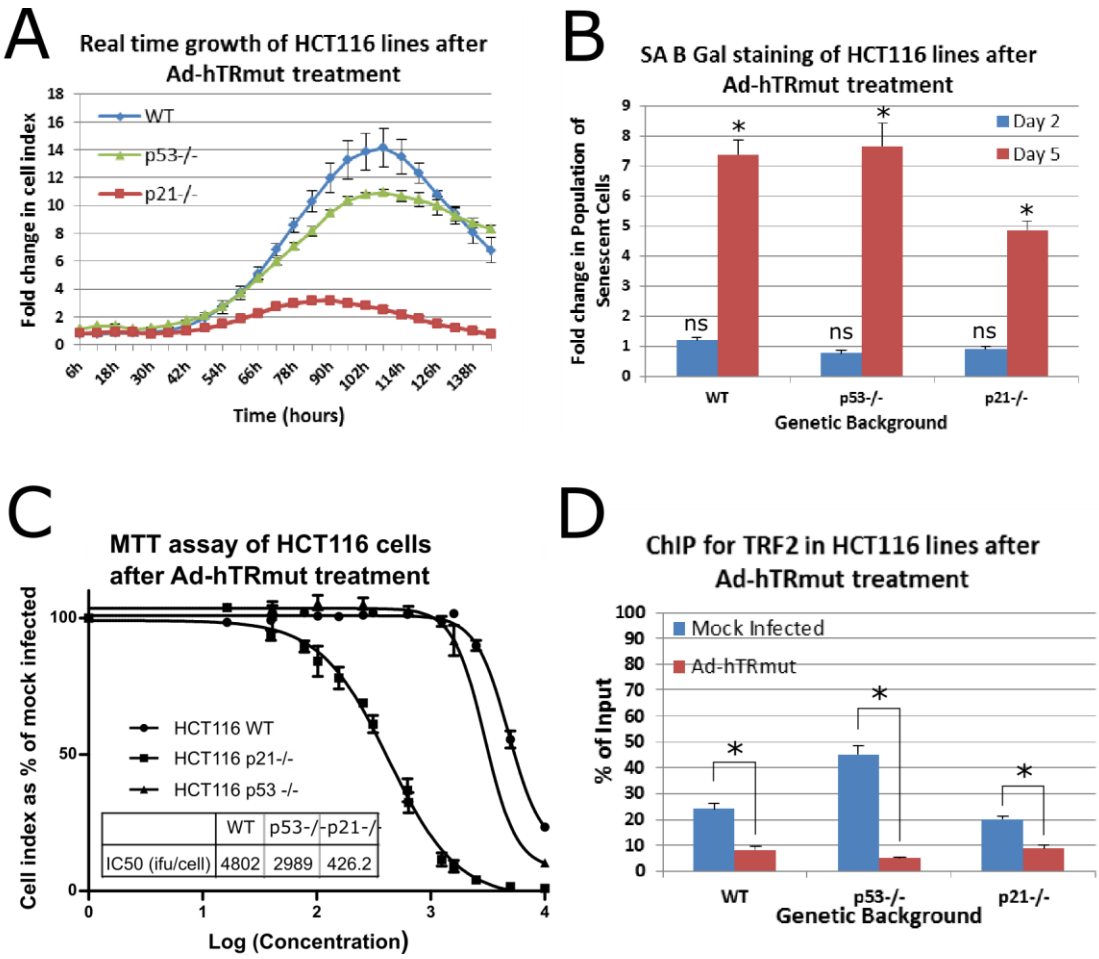


Figure 4.1. Growth of cell lines after treatment with telomere dysfunction triggers.

(A) HCT116 cell lines treated with 5000 ifu/cell of Ad-hTRmut over a short time course starting at 24 post seeding. (B) Senescent phenotype of HCT116 cell lines after treatment with Ad-hTRmut detected by SA B Gal staining. (C) MTT viability assay of cell lines treated with Ad-hTRmut. Cells seeded at 1000 cells/well prior to treatment with viral titration 24h later. MTT assay carried out at 5 days post treatment. Four-parameter dose-response curves were and IC50 values were calculated in GraphPad Prism. (D) ChIP for TRF2 followed by qPCR for telomeric sequences represented as a % relative to the total input to the ChIP. (B, D) p values were calculated by student t-test (*: $p<0.05$, ns: $p>0.05$).

4.2.2. Specific process network profiles are associated with each genetic background

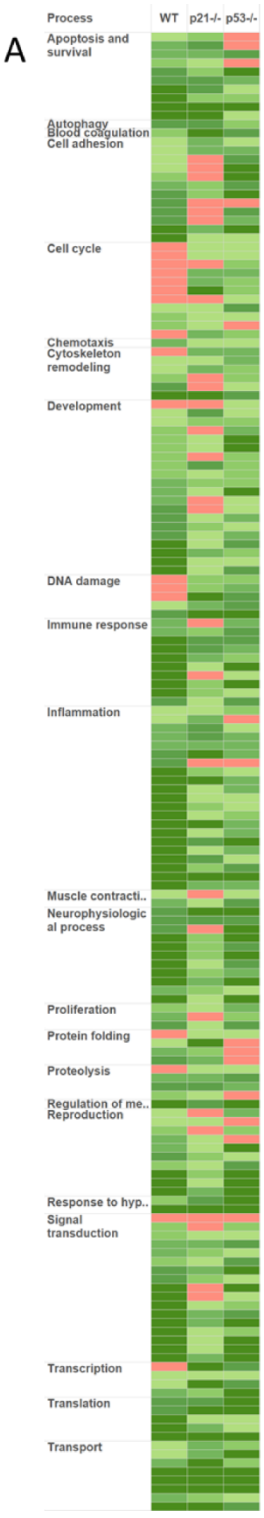
As the $p21^{-/-}$ background was found to be sensitive to telomere uncapping it was important to identify possible biological pathways responsible for the difference in sensitivity between the genetic backgrounds. Cell lines were treated with Ad-hTRmut followed by microarray analysis on Agilent Whole Human Genome (4x44K) arrays. Intra-array normalisation was carried out using the 75th percentile for each microarray. Significant differences in expression between control and treated cells were determined using a paired t-test and Benjamini and Hochberg false discovery rate multiple testing correction of 5%. IDs with a $p<0.005$ were selected for further analysis. The filtered gene lists were then analysed by the process network enrichment analysis methodology (or PNO) described in Chapter 3.

Figure 4.2A shows the overall profile of process networks in the HCT116 lines after Ad-hTRmut induced telomere dysfunction, demonstrating a number of uniquely significant networks in each cell line across a range of biological processes, whilst Figure 4.2B shows a subset of this profile after further filtration to remove processes with few significant networks across the cell lines. WT cells were found to have unique enrichment in cell cycle and DNA damage response process networks. $p21^{-/-}$ had enrichment in cell adhesion networks and $p53^{-/-}$ had enrichment in apoptosis and survival associated networks. In order to further quantify the similarity in the expression profiles a hypergeometric analysis of the process network profile was carried out (Figure 4.2C). This analysis revealed that the WT and $p21^{-/-}$ profiles clustered together whilst the $p53^{-/-}$ profile was on its own branch. This suggests that while the $p21^{-/-}$ is more sensitive to telomere uncapping there is still enrichment in many similar networks to the WT cell lines. However, the $p53^{-/-}$ line has a different profile to the other two lines suggesting activation of different pathways in response to telomere dysfunction.

Further analysis of the networks from the profile in Figure 4.1A highlighted a number of candidate pathways. WT appears to activate a typical DNA damage response. This is highlighted by down regulation of cell cycle components

downstream of the DNA damage signalling components ATM/ATR and Chk1/Chk2 such as Cyclins A, B1, B2 and E. $p53^{-/-}$ cells did not demonstrate significant changes in cell cycle networks but did demonstrate up regulation of death domain receptors and caspases within apoptotic signalling processes, suggesting apoptosis as a method of cell death. $p21^{-/-}$ cells demonstrated expression changes within a range of cell adhesion networks and did not show down regulation in spindle and centromere formation associated genes which were down regulated in other backgrounds.

If additional time was available, further experiments would have been performed to assess the status of the cell lines, i.e. whether the cell lines were being driven towards senescence or apoptosis. For example, in the $p53^{-/-}$ cell line there was enrichment in apoptosis associated pathways, however this was not explored further. It could have been useful to assess the apoptosis status of $p53^{-/-}$ after induction of telomere dysfunction via flow cytometry. However, the focus was on identification of pathways for further analysis and apoptosis was not investigated further.



B

Process	Network	WT	p21-/-	p53-/-
Apoptosis and survival	Death Domain receptors & caspases in apopt..	0.13	0.31	0.00
	Apoptotic nucleus	0.49	0.71	0.04
	Proteolysis in cell cycle and apoptosis	0.25	0.16	0.02
	Anti-Apoptosis mediated by external signals vi..	0.62	0.47	0.79
	Anti-Apoptosis mediated by external signals vi..	0.76	0.39	0.98
	Anti-Apoptosis mediated by external signals vi..	0.79	0.62	0.46
	Apoptosis stimulation by external signals	0.83	0.65	0.16
	Endoplasmic reticulum stress pathway	0.93	0.32	0.24
	Anti-Apoptosis mediated by external signals b..	0.95	0.90	0.96
	Apoptotic mitochondria	0.98	0.83	0.08
Cell adhesion	Synaptic contact	0.10	0.61	0.18
	Platelet-endothelium-leucocyte interactions	0.13	0.57	0.47
	Integrin priming	0.15	0.01	0.62
	Attractive and repulsive receptors	0.21	0.01	0.86
	Integrin-mediated cell-matrix adhesion	0.32	0.00	0.90
	Platelet aggregation	0.48	0.30	0.65
	Cell-matrix interactions	0.68	0.36	0.99
	Amyloid proteins	0.72	0.01	0.01
	Leucocyte chemotaxis	0.77	0.01	0.67
	Cell junctions	0.79	0.00	0.53
	Glycoconjugates	0.95	0.51	0.99
	Cadherins	0.98	0.05	0.10
Cell cycle	Mitosis	0.00	0.12	0.05
	S phase	0.00	0.13	0.18
	G2-M	0.00	0.05	0.41
	Spindle microtubules	0.00	0.46	0.43
	G1-S	0.00	0.52	0.26
	Meiosis	0.00	0.92	0.41
	G1-S Growth factor regulation	0.01	0.01	0.15
	G1-S Interleukin regulation	0.16	0.17	0.75
	G0-G1	0.25	0.22	0.13
	Proteolysis in cell cycle and apoptosis	0.25	0.16	0.02
	Core	0.00	0.49	0.32
DNA damage	Checkpoint	0.00	0.32	0.39
	MMR repair	0.00	0.39	0.55
	DBS repair	0.05	0.93	0.67
	BER-NER repair	0.08	0.46	0.79
	Core	0.76	1.00	1.00

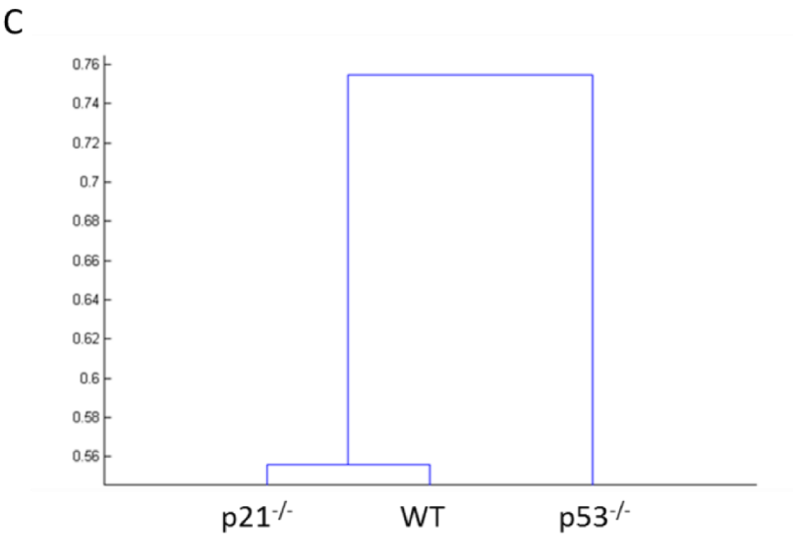


Figure 4.2. Process network enrichment analysis after treatment of HCT116 cell lines with Ad-hTRmut.

(A) Profile view of all three lines post treatment where red rectangles are networks where p <0.05. (B) Profile view after filtering out processes with few/no significant networks. (C) Hypergeometric analysis of figure A quantifying the similarity between the process network profiles.

4.2.3. p21 is required for Cell Cycle and Mismatch Repair regulation

The previous results suggested that Cell Cycle and DNA Damage processes are enriched in the WT background but not in $p53^{-/-}$ or $p21^{-/-}$ backgrounds. To further investigate this, specific networks were chosen for further analysis. Figure 4.3A shows the full network for the cell cycle core process network. In the WT background (data shown) p21 was upregulated and a number of cyclins were down regulated indicating possible shut down of the cell cycle. However, in the $p21^{-/-}$ and $p53^{-/-}$ backgrounds there were no changes in cyclin expression levels. As a proxy, Cdk2 protein levels were also measured (Figure 4.3B). Cdk2 protein levels were reduced in the $p21^{-/-}$ background at day 2 and day 5 but not in the WT and $p53^{-/-}$ backgrounds. This could indicate a p21 independent form of cell cycle arrest in the $p21^{-/-}$ background whilst the WT and $p53^{-/-}$ cells showed no changed in cdk2 protein levels but appeared to down regulate cyclins in a p21 dependent manner. Further to this, there was also a drop in expression of centromere and spindle formation associated genes in the WT background but not in the $p53^{-/-}$ and $p21^{-/-}$ lines.

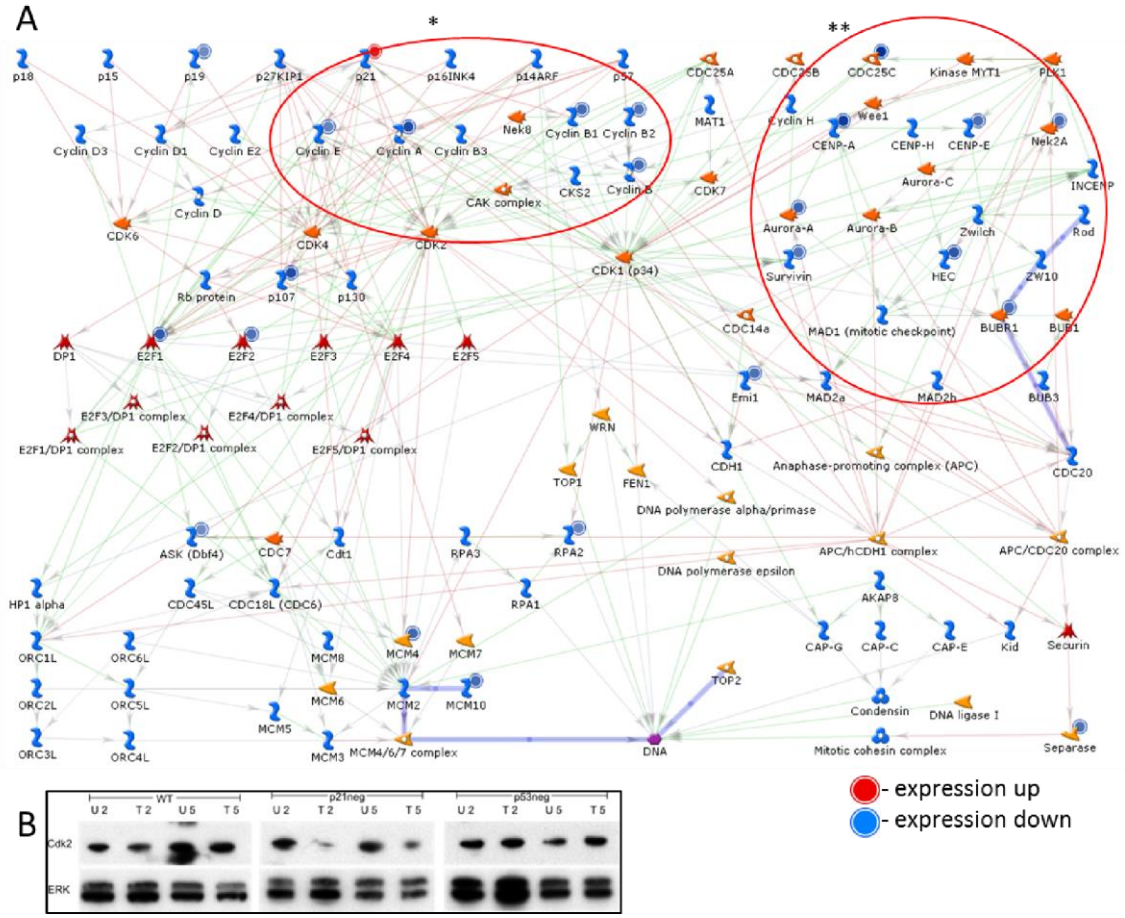


Figure 4.3. Candidate process network: Cell cycle core.

(A) Cell cycle core process network with WT expression data 2 days post hTRmut treatment. * Down regulation of cyclins suggests shut down of the cell cycle in response to telomere dysfunction, ** Down regulation of centromere and spindle formation associated genes. (B) Western blot for Cdk2 with ERK loading controll in all cell lines after 2 or 5 days hTRmut treatment. U = Untreated, T = Treated.

Within the DNA Damage process the mismatch repair process network was enriched in the WT background but not in $p53^{-/-}$ or $p21^{-/-}$. Figure 4.4A shows the full network for the mismatch repair process network. In the WT background (data shown) it can be seen that key MutSα components MSH2 and MSH6 are down regulated. This change was absent in the $p53^{-/-}$ and $p21^{-/-}$ backgrounds. To further investigate this MSH2, MSH6 and MLH1 were measured at the protein level after Ad-hTRmut treatment (Figure 4.4B). Although MLH1 showed no expression changes it was included as it plays an important role in mismatch recognition by the MutSα complex. MSH2 and MSH6 were down regulated in the $p21^{-/-}$ background after treatment but not the other two lines (although the signal for $53^{-/-}$ in MSH2 was poor). MLH1 was unchanged in all lines. This indicates that in the WT background a drop in expression at day 2 does not have an immediate effect on the protein level by day 2 or day 5 whilst in the p21neg background there appears to be loss of the MutSα complex after treatment with hTRmut. This may indicate a role for p21 in the stability of the MutSα complex and implicates the mismatch repair pathway in regulation of telomere dysfunction. The role of the mismatch repair pathway is further investigated in Chapter 5.

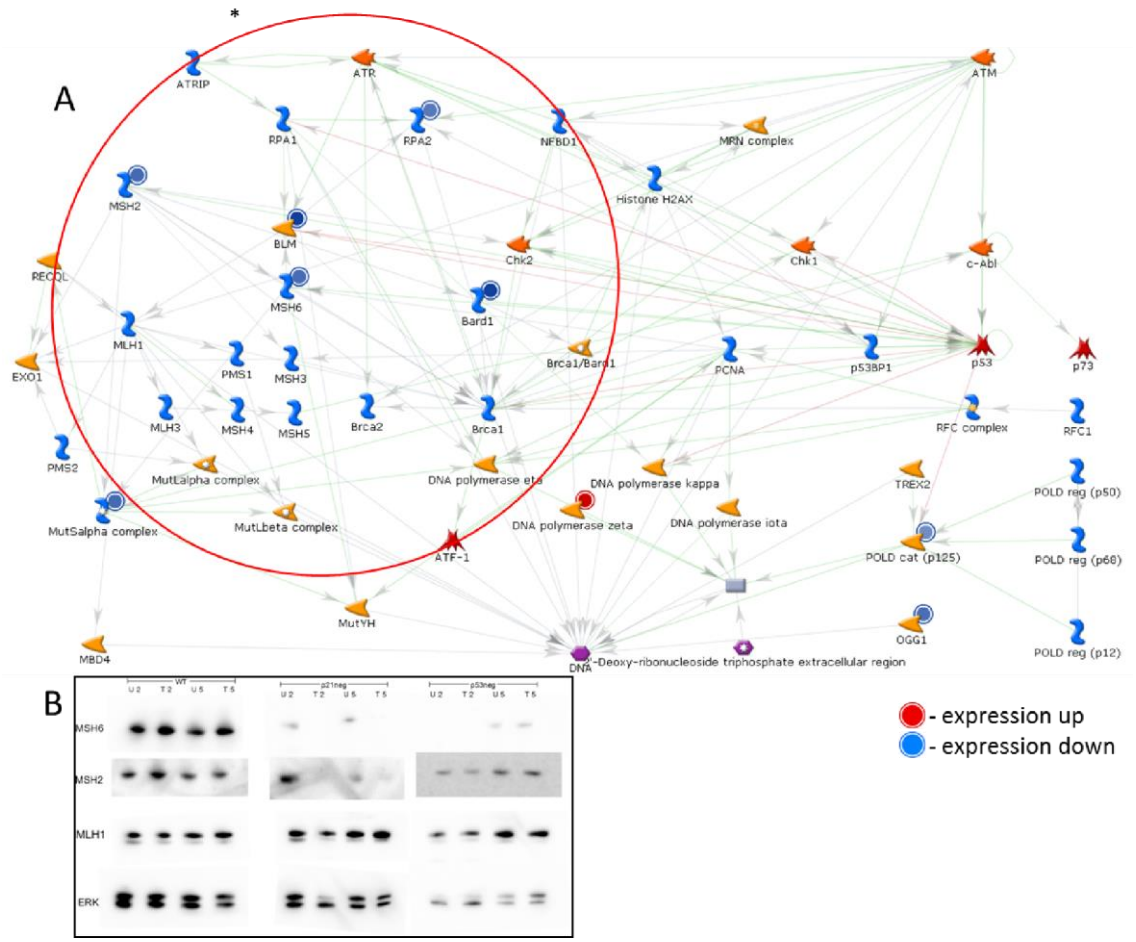


Figure 4.4. Candidate process network: Mismatch Repair.

A. Mismatch repair process network with WT expression data 2 days post hTRmut treatment. * Down regulation of MutS alpha complex components in WT which were unchanged in p21neg and p53neg backgrounds. B. Western blot for MSH6, MSH2 and MLH1 with ERK loading control in all cell lines after 2 or 5 days hTRmut treatment. U = Untreated, T = Treated.

4.2.4. Transient upregulation of hTERT may be associated with resistance to telomere dysfunction

It was also hypothesised that differences in sensitivity between the cell lines may be due to differences in telomerase regulation. A pilot experiment was carried out using a hTERT-luciferase reporter plasmid to measure hTERT promoter activity in response to Ad-hTRmut (Figure 4.5A). hTERT promoter activity was increased 4 fold in the WT background after 2 days of Ad-hTRmut treatment but was not significantly up regulated in the other cell lines. WT was found to be significantly upregulated when directly compared with the $p21^{-/-}$ and $p53^{-/-}$ ($p<0.05$) whereas there was no significant different when comparing $p21^{-/-}$ with $p53^{-/-}$ ($p>0.05$). In order to investigate this further hTERT mRNA expression levels were directly measured by qPCR 2 or 5 days after hTRmut treatment (Figure 4.5B). At day 2, hTERT mRNA had decreased significantly in the WT and $p53^{-/-}$ backgrounds but was unchanged in $p21^{-/-}$. By day 5, hTERT mRNA levels were elevated 4 and 2 fold in WT and $p53^{-/-}$ lines respectively. In $p21^{-/-}$ cells hTERT appeared to be elevated by day 5, however the result was not significant. At day 5, WT was found to be significantly upregulated when directly compared with the $p21^{-/-}$ and $p53^{-/-}$ ($p<0.05$) whereas there was no significant different when comparing $p21^{-/-}$ with $p53^{-/-}$ ($p>0.05$). This indicates that telomere uncapping induces hTERT expression in WT and $p53^{-/-}$ backgrounds and $p21^{-/-}$ may be deficient in this response.

Changes in expression of the RNA component of telomerase, hTR, can also lead to changes in telomerase activity (Feng et al., 1995). Therefore hTR expression levels after Ad-hTRmut treatment was also measured (Figure 4.5C). However, there were no significant changes in expression in hTR in any of the cell lines at either time point. Therefore, any role of telomerase is likely to be linked to hTERT rather than hTR. The next logical step was to assay telomerase activity in response to telomere uncapping (Figure 4.5D). Telomerase activity was measured by TRAP assay after Ad-hTRmut treatment of HCT116 cell lines. At day 2, WT cells had significant upregulation of telomerase activity, however the effect was very small with only a 1.2 fold increase in activity. $p53^{-/-}$ and $p21^{-/-}$ did not show any change in telomerase activity at Day 2. However, by Day 5 telomerase activity had decreased significantly in all cell lines, dropping to 0.6

fold in WT and p53$^{-/-}$ and to 0.4 fold in p21$^{-/-}$. These results indicate that hTERT up regulation may be involved in regulation of telomere dysfunction independently of telomerase activity and that this effect is p21 dependent.

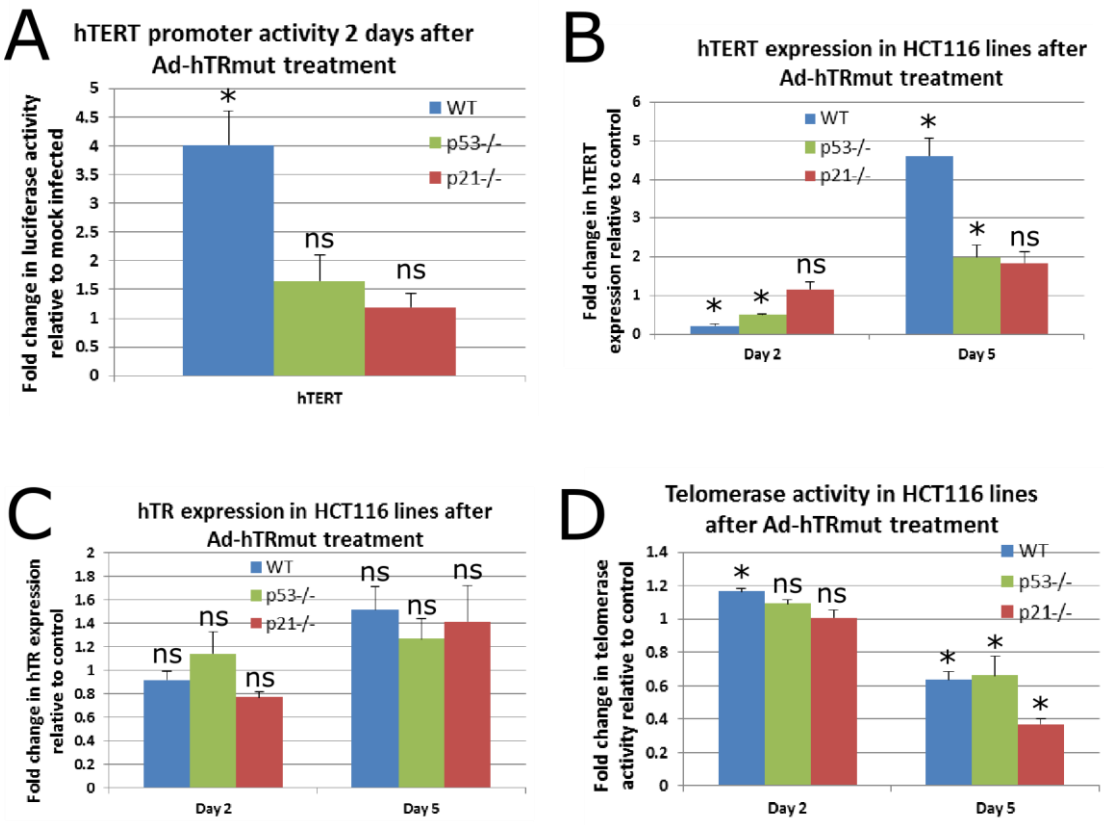


Figure 4.5. Telomerase regulation in response to Ad-hTRmut treatment.

(A) Luciferase assay for hTERT promoter activity 2 days after hTRmut treatment represented as fold change relative to control. (B) qPCR measurement of hTERT mRNA after hTRmut treatment represented as fold change relative to control. (C) qPCR measurement of hTR mRNA after hTRmut treatment represented as fold change relative to control. (D) TRAP assay for telomerase activity after hTRmut treatment represented as fold change relative to control. (A,B,C and D) p values were calculated by student t-test (*: $p<0.05$, ns: $p>0.05$).

4.3. Discussion

4.3.1. Genetic background affects sensitivity to telomere uncapping

p21 and p53 are key in tumour suppression and are points of convergence in the DNA damage response pathway along with p16/pRB (Salama et al., 2014). Severing of the DNA damage signalling pathway has been previously shown in taz1 deficient yeast which are unable to initiate a telomere dysfunction induced DNA damage response (Carneiro et al., 2010). Therefore it was thought that $p21^{-/-}$ and $p53^{-/-}$ cell lines may be resistant to telomere dysfunction due to inability to signal dysfunctional telomeres. Indeed, PNO analysis revealed that p21 and p53 deficient cell lines had no enrichment in DNA damage response pathways indicating deficiency in DNA damage signalling. However, induction of rapid telomere uncapping by Ad-hTRmut had a clear difference in rate of growth reduction in the WT, $p21^{-/-}$ and $p53^{-/-}$ backgrounds. This was first shown through a real time growth assay and corroborated with a secondary MTT viability assay which found $p21^{-/-}$ cells to be particularly sensitive to telomere uncapping. Further to this, analysis of the senescence phenotype highlighted that $p21^{-/-}$ cells had less SA ß Gal staining relative to WT and $p53^{-/-}$ cells after induction of telomere dysfunction. $p21^{-/-}$ cells did have a significant increase in senescent, however the difference may indicate an alternate route to cell death in a population $p21^{-/-}$ of the cells.

p53 deficiency did not seem to affect the cells response to telomere dysfunction as $p53^{-/-}$ cells had an almost identical growth phenotype to the WT cells. p53 independent senescence has been previously observed. It has been shown that $p16^{Ink4a}$ may play a role in senescence induction in p53 deficient cells and that senescence could be bypassed by ablation of $p16^{Ink4a}$ (Jacobs and de Lange, 2004). p21 has also been shown to be regulated independently of p53 by Chk2 (Aliouat-Denis et al., 2005). Therefore in the $p53^{-/-}$ cell line p53 deficiency may be compensated for by other pathways such as the p16/pRB senescence pathway or p21 up regulation by other mechanisms bypassing p53 entirely. PNO analysis revealed up regulation of apoptosis associated genes which were not changed in

the WT background as further evidence of rerouted signalling in the absence of p53.

In addition to the increased sensitivity to uncapping, $p21^{-/-}$ cells also demonstrated an inability to down regulate spindle and centromere formation genes which may point to cell death via another mechanism. Telomere uncapping has been linked to mitotic stress. While this was initially linked to p53, it appears to be reliant on the p53/p21 pathway indicating that p21 is a key effector of the mechanism (Thanasoula et al., 2010; Thanasoula et al., 2012b). Therefore it seems likely that in $p21^{-/-}$ progression through the cell cycle occurs uninhibited. This could lead to mitotic stress in the cells and therefore mitotic catastrophe as a mechanism of cell death which does not occur in the other backgrounds.

After induction of telomere dysfunction with Ad-hTRmut, hTERT expression was transiently down regulated 2 days post treatment and was subsequently up regulated by day 5 in the WT and $p53^{-/-}$ backgrounds. However, this did not correspond with a change in telomerase activity in any of the cell lines. However, hTERT has been shown to have roles independent of telomerase. For example, it has recently been shown to have anti-apoptotic effects in cells under oxidative stress (Kida et al., 2013). It is possible that hTERT is able to mask the telomere in some way in the WT background leading to a disruption in the typical signalling response. Reduced hTERT up regulation in $p53^{-/-}$ and $p21^{-/-}$ backgrounds could play a role in their sensitivity if this is the case, particularly in $p21^{-/-}$ which demonstrated no hTERT up regulation. Therefore, hTERT may also play a wider regulatory role than previously thought in the context of telomere dysfunction and warrants further investigation.

Chapter 5: Mismatch Repair and Telomere dysfunction

5.1. Introduction

During DNA synthesis, polymerases usually operate with very high accuracy. While they do have a low error rate they are not perfect and during replication some sequences can cause slippage, such as sequences with high repetition, which can increase the error rate (Jun et al., 2006). The mismatch repair (MMR) system primarily operates as a system which can detect lesions caused by DNA polymerase slippage and then remove the error. DNA polymerase can then reattempt DNA synthesis and correct the error (Jun et al., 2006). The MMR system can recognise small loops within DNA, which occur due to mismatched base pairs or by insertion/deletion loops (IDLs), and repair them (Harfe and Jinks-Robertson, 2000; Hsieh, 2001; Kolodner and Marsischky, 1999; Michailidi et al., 2012). The MMR system is essential in DNA homeostasis. Loss of this repair mechanism leads to a 100-1000 fold increase in DNA replication error rate leading to increased disposition to cancer (Burgart, 2005; Jascur and Boland, 2006; Vilar and Gruber, 2010). Also the MMR system has been implicated in telomere function and aging (Conde-Perezprina et al., 2012).

The MMR system is a highly conserved system from bacteria to humans and involves excision and resynthesis of DNA. The process can be divided into four steps. First, the mismatch is recognised in the DNA by MutS proteins. Second, enzymes are recruited to the lesion which can repair the mismatched DNA. Third, the mismatched base/bases are excised. Finally the DNA is resynthesized by DNA polymerase (Jun et al., 2006). In eukaryotes, recognition of DNA lesions is accomplished by two heterodimers composed of MSH proteins (MutS homologs), MutSα (MSH2-MSH6) and MutSß (MSH2-MSH3). While the core MMR proteins are highly conserved, eukaryotic MMR also incorporates a number of accessory proteins with critical roles which vary depending on the nature of mismatch being repaired (Kunkel and Erie, 2005). MutSα primarily recognises single base-base and IDL mismatches whilst MutSß is responsible for recognition of IDLs containing short stretches of nucleotides (McCulloch et al., 2003). After a

lesion is detected MutLα (MutL homolog) is recruited (MLH1-PMS2 heterodimer in humans) and is used to repair a wide variety of mismatches whilst greatly enhancing the efficiency of mismatch recognition by MutS complexes after binding (Vaish and Mittal, 2002). The MutS complexes have also been found to interact with proliferating cell nuclear antigen (PCNA) which plays a role in discriminating between strands in the DNA duplex and correctly positioning the MMR system at the mismatches base(s) (Flores-Rozas et al., 2000; Kunkel and Erie, 2005; Lau and Kolodner, 2003; Vaish and Mittal, 2002). After recognition the mismatched base is excised by ExoI which is aided through physical interactions with MutS and MutL complexes (Tran et al., 2001). PCNA is also known to interact with ExoI and to aid in the excision process and eventual DNA resynthesis (Kunkel and Erie, 2005).

Although the primary role of the MMR system is to repair DNA mismatches during replication, the system has implicated roles in a number of other biological processes and signalling cascades. MMR has been found to play an important role in meiotic and mitotic recombination, DNA damage signalling and various aspects of DNA metabolism (Jiricny, 2006). The MMR system does appear to have a regulatory role at the telomere outside of mismatch repair, however the specifics are not yet clear. Previously MMR deficiency was shown to increase survival and have anti-ageing effects in telomerase negative models, negating the effects of critically short telomeres in either PMS2 (MutL complex) or MSH2 (MutSα complex) deficient mice (Martinez et al., 2009; Siegl-Cachedenier et al., 2007). However, more recently in a number of telomerase positive tumour derived cell lines, MMR deficiency was shown to increase the rate of telomere shortening indicating that the MMR system is also required to maintain telomere integrity (Mendez-Bermudez and Royle, 2011). It could be that differences in telomere status have a significant effect on the outcome of MMR deficiency and therefore further characterisation of the pathway in the context of telomere dysfunction is required.

In Chapter 4 a number of candidate pathways which may be important in regulation of telomere dysfunction were identified including the mismatch repair pathway. WT cells engaged the MMR pathway whilst p21 and p53 negative cells

did not. It was hypothesised that mismatch repair may either play a role in mediating telomere damage after uncapping of the telomeres by Ad-hTRmut or that the mismatch repair pathway may have a role in signalling leading to regulation of other pathways. Therefore, it was desirable to further investigate the mismatch repair pathway and characterise its role in telomere regulation and response to telomere uncapping.

5.2. Results

5.2.1. MSH2 plays a role in mediating telomere dysfunction

In Chapter 4 key components of the MutSα complex, MSH2 and MSH6, were down regulated at the protein level in the $p21^{-/-}$ background after induction of telomere dysfunction by Ad-hTRmut but were unchanged in the WT background. However, in all backgrounds MSH2 and MSH6 were down regulated at the mRNA level. Therefore, further investigation was required to understand these conflicting results. In order to investigate whether the mismatch repair process was essential to regulation of telomere dysfunction key components, MLH1 and MSH2, were knocked down by RNAi followed by 2 days of Ad-hTRmut treatment. PNO analysis was then carried out to determine the effects of interruption of the mismatch repair pathway. Figure 5.2A demonstrates the full profile of process networks comparing WT, $p21^{-/-}$, $p53^{-/-}$ and WT+MSH2 siRNA profiles. After telomere uncapping WT had significant enrichment of specific DNA damage process networks whilst $p21^{-/-}$ and $p53^{-/-}$ cells did not. This is indicative of a DNA damage response to telomere uncapping in WT cells which is absent in $p21^{-/-}$ and $p53^{-/-}$ cell lines. Knockdown of MSH2 in WT cells prior to Ad-hTRmut treatment resulted in a shift in process network profile with apparent similarity to $p21^{-/-}$ cells after Ad-hTRmut treatment with shared networks in cell adhesion and cytoskeleton remodelling. These co-treated cells did not demonstrate enrichment of DNA damage pathways indicating that MSH2 may be involved in the DNA damage signalling process.

The visual interpretation of the process network profile was quantitatively confirmed by a clustering analysis based on hypergeometric probability of

overlap in the profiles (Figure 5.1B). Clustering revealed that the WT process network profile shifts to be more similar to $p21^{-/-}$ after Ad-hTRmut treatments. This indicates that $p21^{-/-}$ cells may be deficient in mismatch repair activity in response to telomere dysfunction. To further assess the effect of mismatch repair interruption on cell viability an MTT assay was carried out. HCT116 cell lines were transfected with siRNA for MLH1 or MSH2 followed by 2 days of Ad-hTRmut treatment. Figure 5.1C demonstrates cell viability after knockdown of MMR targets followed by Ad-hTRmut treatment. In all lines sensitivity to telomere uncapping was enhanced relative to scrambled siRNA control after knockdown of both MLH1 and MSH2, indicating that MMR activity may be important in response to telomere dysfunction.

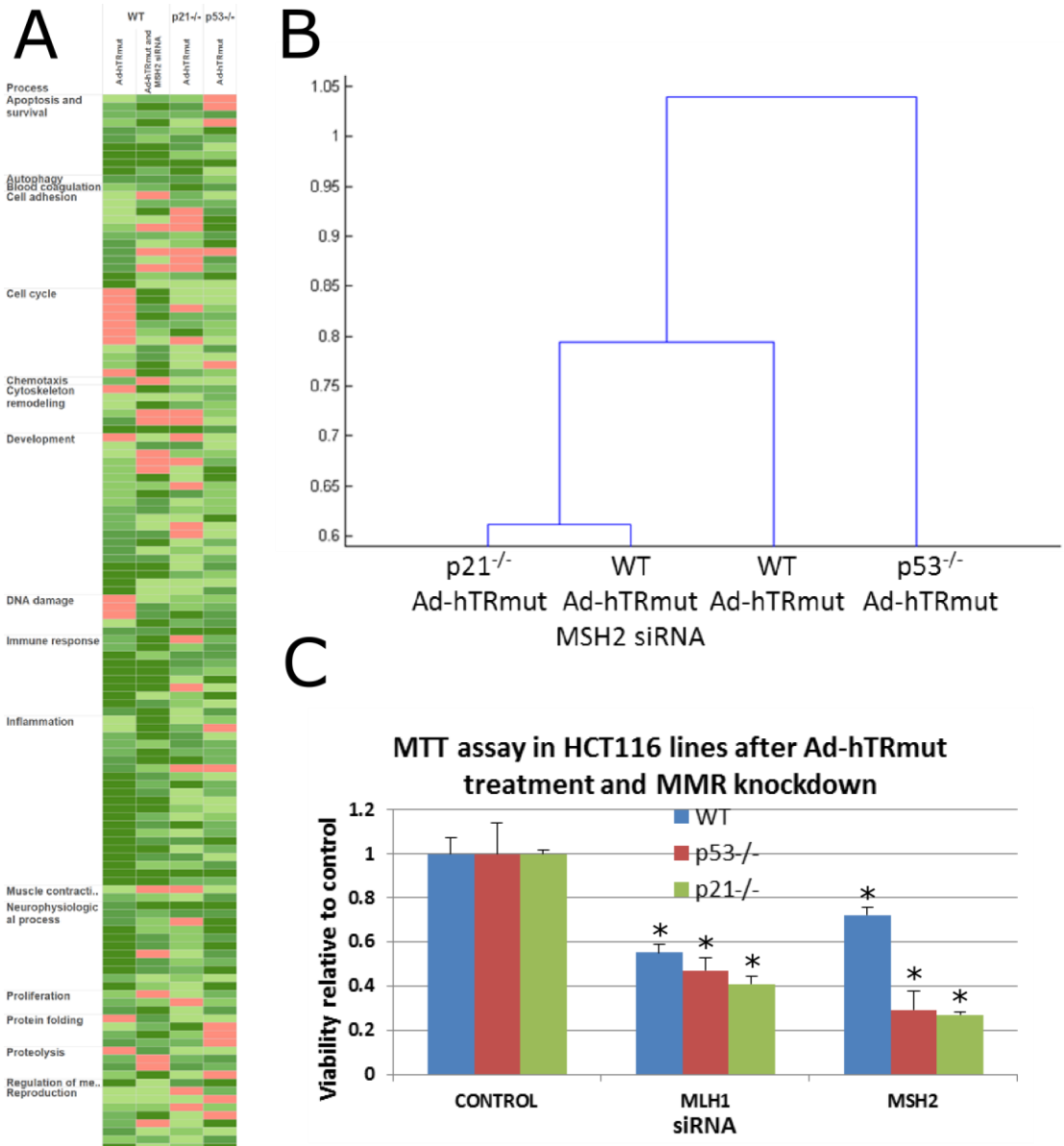


Figure 5.1. Knockdown of MSH2 in HCT116 WT enhances sensitivity to telomere uncapping.

(A) Process network enrichment profile comparing WT, p21neg, p53neg and WT+MSH2 siRNA using same array data as in figure 4.2. (B) Hypergeometric analysis of the profile in A. (C) MTT viability assay of HCT116 lines after knockdown of either MLH1 or MSH2 followed by 2 days of hTRmut treatment. Control is hTRmut treatment only. P values were calculated using student t-test (*: p>0.05).

5.2.2. p21 and p53 are involved in a MutSα regulatory network

In order to further investigate regulatory networks linking p21 to regulation of the MutSα complex network building was carried out in MetaCore. p21, MSH2 and MSH6 were entered as a manual list in the MetaCore "build network from list" section. The Auto expand network building algorithm was then run under default settings (50 network nodes and "use canonical pathways" selected). This algorithm then builds a network using the input list as a starting point and tries to connect entities in the list in 2 or less steps based on known interactions in the literature. The network was then reduced to only include nodes which were connected to all other nodes in the network in 2 steps or less. This resulted in the network shown in Figure 5.2A. The network produced includes p53, p21, cMyc and the MutSα complex (which contains MSH2 and MSH6) with a range of known interactions. p21 is shown in a reciprocal deactivating interaction with cMyc. However, it also reciprocally activates p300 which in turn induces transcription of cMyc, p53 and MutSα. Also, cMyc and p53 both induce expression of MutSα whilst p53 also induces expression of p21. The result appears to be a pair of positive feedback loops initiated by activation of p300 by p21. In the WT background the network is uninterrupted leading to expression of cMyc and the MutSα complex. In the $p53^{-/-}$ background the p21, p300, p53 loop would be interrupted however the p21, p300 activating loop would still be present, allowing for downstream activation of cMyc and MutSα. In the $p21^{-/-}$ background, if p21 is central to the signalling cascade, then neither of the loops would be activated resulting in reduced cMyc and MutSα expression.

To confirm whether telomere uncapping by Ad-hTRmut does indeed involve regulation of c-Myc expression and whether this is differentially affected in $p21^{-/-}$ and $p53^{-/-}$ backgrounds, each cell line was transfected with a c-Myc promoter luciferase reporter vector. 24h post-transfection, cells were infected with Ad-hTRMut and incubated for an additional 48h prior to harvest and luciferase assay (Figure 5.2B). It was found that cMyc is upregulated in all of the cell lines following Ad-hTRmut treatment. However, genetic background did affect the level of this upregulation with the $p21^{-/-}$ demonstrating a 2 fold increase in cMyc promoter activity whilst the WT and $p53^{-/-}$ demonstrated up to 8 and 4 fold

increases respectively. This supports the pathway presented in figure 5.2A in which WT has full activation of the pathway, $p53^{-/-}$ is only partially interrupted and $p21^{-/-}$ would be largely ablated.

In order to validate this model focus was directed to WT and $p21^{-/-}$ cell lines. The cell lines were treated with either Ad-hTRmut, MSH2 siRNA transfection or a co-treatment where cells were treated with Ad-hTRmut after MSH2 siRNA transfection followed by qPCR measurement of p21 or p300 mRNA levels (Figure 5.2C). At day 2 Ad-hTRmut treatment p300 expression greatly increased in both cell lines. In WT this was followed by increased p21 expression by day 5 of treatment which was not present in the $p21^{-/-}$ background as expected. siRNA knockdown of MSH2 followed by Ad-hTRmut treatment ablated expression of both p300 and p21 in WT and p300 in the $p21^{-/-}$ cells. These results indicate that MSH2 is required for downstream activation of p300 followed by p21 after telomere uncapping.

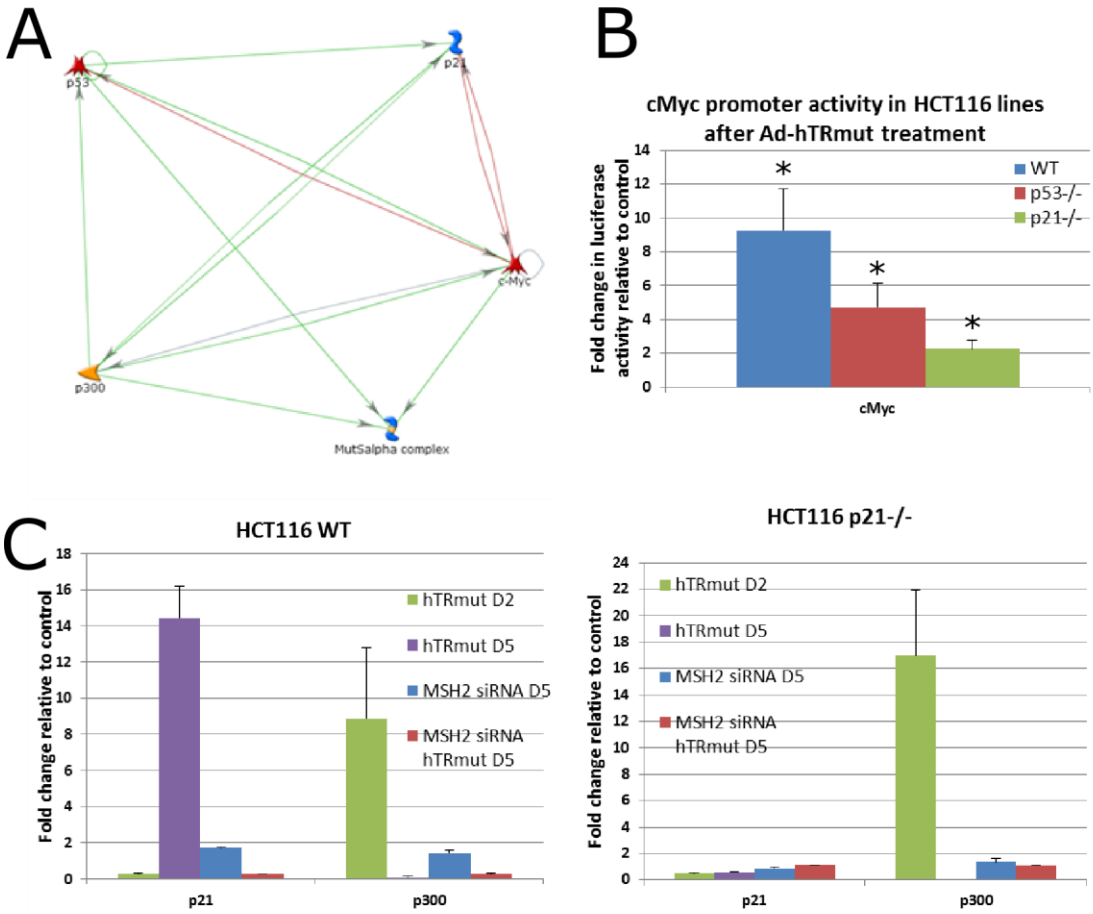


Figure 5.2. Network building and validation reveals an interaction between p21, p53, p300, cMyc and the MutSα complex.

(A) Network resulting from seed nodes p21, MSH2 and MSH6. The network is a subset highlighting interactions between 5 specific genes (B) cMyc promoter luciferase reporter activity after 2 days of hTRmut treatment in HCT116 lines. (C) qPCR expression analysis of p21 and p300 in WT and p21-/- cells after Ad-hTRmut treatment at day 2, Ad-hTRmut treatment at day 5, MSH2 siRNA only and Ad-hTRmut/MSH2 siRNA co-treatment.

5.3. Discussion

In the process network analysis the Mismatch Repair process was found to be significantly enriched in WT background but not the p53 or $p21^{-/-}$. Further investigation was needed to understand what role the mismatch repair process plays in regulation telomere dysfunction. MutSα components MSH2 and MSH6 were absent after Ad-hTRmut treatment in the $p21^{-/-}$ background. In addition, knockdown of MSH2 and MLH1 resulted in enhanced sensitivity to telomere uncapping in all genetic backgrounds whilst MSH2 knockdown also caused a shift in the overall WT expression profile. Furthermore, a regulatory link between p21, p53, cMyc, p300 and the MutSα complex was suggested via network building in MetaCore. This network revealed two positive feedback loops which can both be initiated by p21 and result in the activation of the MutSα complex. Further expression analysis of the pathway indicated that p300 expression increased followed by p21 expression at a later timepoint, whilst MSH2 knockdown ablated these expression changes completely. Together, these findings suggest a role for the mismatch repair process in response to telomere dysfunction and that p300 may play a central role in regulation of the overall signalling pathway.

In addition to the results shown here, at the protein level, p21 has been shown to inhibit the cMyc-Max formation (Kitaura et al., 2000). It has also been shown that Max can bind to the MSH2 protein, though the function of this is unclear (Mac Partlin et al., 2003). One possibility is that Max helps with stabilisation of the MSH2 protein and that in the $p21^{-/-}$ background this does not occur, as Max will be bound by cMyc instead. Therefore, in the WT background MSH2 is stabilised which may explain the persistence of MSH2 protein even when the expression level was reduced. Also, MSH2 and MSH6 have been shown to be degraded by the ubiquitin-proteasome pathway in a cell type dependent manner which could account for the reduction in MSH2 and MSH6 levels in the $p21^{-/-}$ background after treatment (Hernandez-Pigeon et al., 2004).

Another route to cell death in the $p21^{-/-}$ background could be an inability to slow its cell cycle in response to telomere dysfunction. It was found that the WT cell line had down regulation of central cell cycle genes such as the cyclins which

regulate progression through the cell cycle whereas the $p21^{-/-}$ background demonstrated no expression changes in these genes. Also p21 is known to negatively regulate the cell cycle via inhibition of CDK/Cyclin complexes (Harper et al., 1993). It is possible that in the $p21^{-/-}$ background damage accumulates and inability to slow the cell cycle and mount a DNA damage response, including the mismatch repair pathway, causes the $p21^{-/-}$ cells to progress through the cell cycle unchecked and die rapidly due to accumulation of DNA damage such as telomeric fusions. This could lead cell death through an alternative mechanism such as mitotic catastrophe which has been associated with telomere dysfunction (Hampel et al., 2013).

Overall there is evidence that p21 is required for activation of the mismatch repair pathway and this pathway is essential in cellular response to telomere dysfunction. The data presented here suggests that this occurs through a regulatory process that occurs at both the transcriptional and protein levels. At the transcriptional stage, p21 appears to interact with p300 which induces both cMyc and MSH2 expression (Wang et al., 2011a; Wang et al., 2007) whilst cMyc protein also reinforces MSH2 expression (Menssen and Hermeking, 2002). At the protein level, p21 interacts with cMyc, inhibiting the formation of cMyc-Max complex. This may lead to cMyc protein being free to further induce MSH2 expression whilst enabling Max to bind MSH2 protein, potentially in a stabilising role. I hypothesise that p21 is required for signalling and further activation of the mismatch repair pathway in response to telomere dysfunction and that cellular survival is mediated as a result of this process.

The model presented above offers a number of opportunities for further investigation. It is worth noting that p21 is not a common mutation associated with oncogenesis. However, p21 is involved in a wide range of biological processes and has been shown to be involved in regulation of the cell cycle, regulation of gene transcription, modulation of apoptosis and DNA repair (Abbas and Dutta, 2009) which will all have clinical relevance. The model outlined here indicates that p21 may be central to the regulation of mismatch repair in response to telomere dysfunction. Therefore, clinically, it may be relevant to assess the p21 status of a cancer prior to use of telomere based therapeutic

strategies. However, further work is required to validate this model. Firstly, the interactions outlined here require further experimental validation. In particular the interaction between p21, p300 and MutSalpha. p300 is known to be a transcriptional activator of p21 (Abbas and Dutta, 2009) however its role in activation of mismatch repair pathways in response to telomere dysfunction is not well characterised. Experimental demonstration of the binding of p300 to the MSH2 promoter would confirm whether MSH2 is a transcriptional target of p300. Also, the link between p21 and mismatch repair requires further proof. One method could be to knock out p21 in the WT cell line and then induce telomere dysfunction. If mismatch repair is activated via p21 in response to telomere dysfunction then we would expect to see loss of regulation of mismatch repair when p21 is knocked down. As p300 is central to the model, knockdown of p300 would be expected to have the same outcome. In addition, the function and role of mismatch repair during telomere dysfunction is not clear. First, I would like to be able to demonstrate that proteins such as MSH2 become localised to the telomeres. This would inform whether MSH2 has a direct effect at the telomere. This could be coupled with a mismatch repair deficient model cell line. If mismatch repair is required to alleviate stress during telomere dysfunction then one would expect ablation of MSH2 to have a similar affect as p21 and/or p300 knockdown. The method of death after induction of telomere dysfunction was also not fully explored. It was hypothesised that $p21^{-/-}$ cells may reach cell death through another mechanism. Given the opportunity, experiments to assess the apoptosis and senescence status of the cells would be performed after knockdown of MSH2. It may be that ablation of mismatch repair causes a shift from cellular senescence to apoptosis, or vice versa, after induction of telomere dysfunction. In addition, assays could be performed to investigate the cell cycle stage at which these cells are dying or ceasing to proliferate. It was thought that $p21^{-/-}$ cells were not slowing their cell cycle and that a build-up of damage was occurring. This could be assayed by flow cytometry to determine if there is a shift in cell cycle status compared to the WT. In addition, it could be informative to look for other forms of DNA aberration such as chromosomal fusions which may occur if the cell cycle continues inhibited when uncapped telomeres are present.

Chapter 6: The effects of telomerase inhibition can be enhanced by hypoglycaemia

6.1. Introduction

In Chapters 4 and 5 we used Ad-hTRmut as a trigger which rapidly induces telomere dysfunction by uncapping the telomeres. This leads to a reduced rate of cell growth, reduced viability and induces senescence. However, under natural settings the telomeres shorten slowly as a result of the end replication problem during cell division. This can also be accelerated by oxidative damage and other events (Cairney et al., 2012). In order to model this method of telomere shortening the telomerase inhibitor GRN163L (Imetelstat) was introduced. GRN163L is a lipid-conjugated 13-mer oligonucleotide sequence that is complementary to and binds with high affinity to the RNA template of telomerase, thereby directly inhibiting telomerase activity. The compound has a proprietary thio-phosphoramidate backbone, which is designed to provide resistance to the effect of cellular nucleases, thus conferring improved stability in plasma and tissues, as well as significantly improved binding affinity to its target. The lipid group also allows the compound to permeate cellular membranes. The tissue half-life of GRN163L, or the time it takes for the concentration or amount of GRN163L to be reduced by half, in bone marrow, spleen, liver and tumour has been estimated to be 41 hours in humans, based on data from animal studies and clinical trials. Inhibition of telomerase by GRN163L in proliferating cells leads to telomere shortening and eventually replicative senescence. However, there are also environmental factors which can lead to telomere shortening. For example, work within our lab identified hypoglycaemia to be a negative regulator of telomerase activity which also led to telomere shortening. In addition to this, glucose restriction has been shown to enhance the effects of a telomerase inhibitor in breast cancer and reduce the lifespan of telomere dysfunctional mice (Missios et al., 2014; Wardi et al., 2014).

In order to improve our understanding of telomere dysfunction it is important to compare a range of triggers. Just as genetic background resulted in different pathways being activated in Chapter 4, different triggers may induce different

signalling processes and downstream pathways. Characterising these pathways will help us understand how cancers can repair or tolerate telomere dysfunction and also mechanisms of oncogenesis. The aim of this chapter is to investigate the effects of alternative telomere dysfunction triggers and synergistic effects between them. The first objective was to investigate the phenotypic effect of telomerase inhibition by GRN163L. The next objective was to compare the effects of telomerase inhibition with telomere uncapping using the PNO analysis method. Finally the effects of telomerase inhibition under hypoglycaemic conditions were investigated in order to identify possible pathways and mechanisms by which the cells could be sensitised to telomerase inhibition.

6.2. Results

6.2.1. HCT116 cell lines are resistant to telomerase inhibition

In order to investigate the effects of telomerase inhibition, a long term culture was set up with HCT116 WT and $p21^{-/-}$ cells cultured in media containing 5 μM GRN163L or Mismatch control. Cells were counted and reseeded on Monday and Friday each week. 3h post seeding the cells were treated with either GRN163L or Mismatch control. Cells were also re-treated every Wednesday with fresh media containing GRN163 or Mismatch control. Over the time course, long term telomerase inhibition caused a very mild reduction in cumulative population doublings of the cell lines (Figure 6.1A). By Day 245, GRN163L treated WT cells were reduced by 20 population doublings relative to mismatch control while $p21^{-/-}$ cells demonstrated mildly increased sensitivity and were reduced by 35 population doublings at the same time point. However, in both cases the cultures continued to grow after over 250 days of sustained telomerase inhibition.

In order to identify a cause for the reduction in population doublings the senescent population of cells was measured by SA ß Gal assay (Figure 6.1B). After long term telomerase inhibition by GRN163L, both WT and $p21^{-/-}$ cells demonstrated an increase in SA ß Gal stained cells. In WT the population of senescent cells in GRN163L treated cultures ranged from 2 fold to 5 fold over the control while in $p21^{-/-}$ the population of senescent cells ranged from 5 fold to 9 fold over the control. This contrasted with Ad-hTRmut induced telomere dysfunction (Chapter 4) where the $p21^{-/-}$ cells demonstrated a decreased SA ß Gal staining relative to WT cells after telomere uncapping. This may indicate differences in regulatory mechanisms between cell lines depending on the method of inducing telomere dysfunction. To confirm that the effects seen were due to telomerase inhibition telomerase activity over the time course was measured by a TRAP assay and telomere length was measured by telomere restriction fragment southern blot (Figure 6.1C and Figure 6.1D). Telomere restriction fragment southern blot was carried out by Sharon Burns, University of Glasgow. In both cell lines telomerase activity was reduced to below 10% after

GRN163L treatment. In addition, by day 185, telomere shortening had occurred in both WT and p21$^{-/-}$ backgrounds. Therefore, this indicates that Telomerase was inhibited over the time course resulting in telomere shortening. However, it is not clear if the effects on cell doubling rate were due to telomeres reaching critical length.

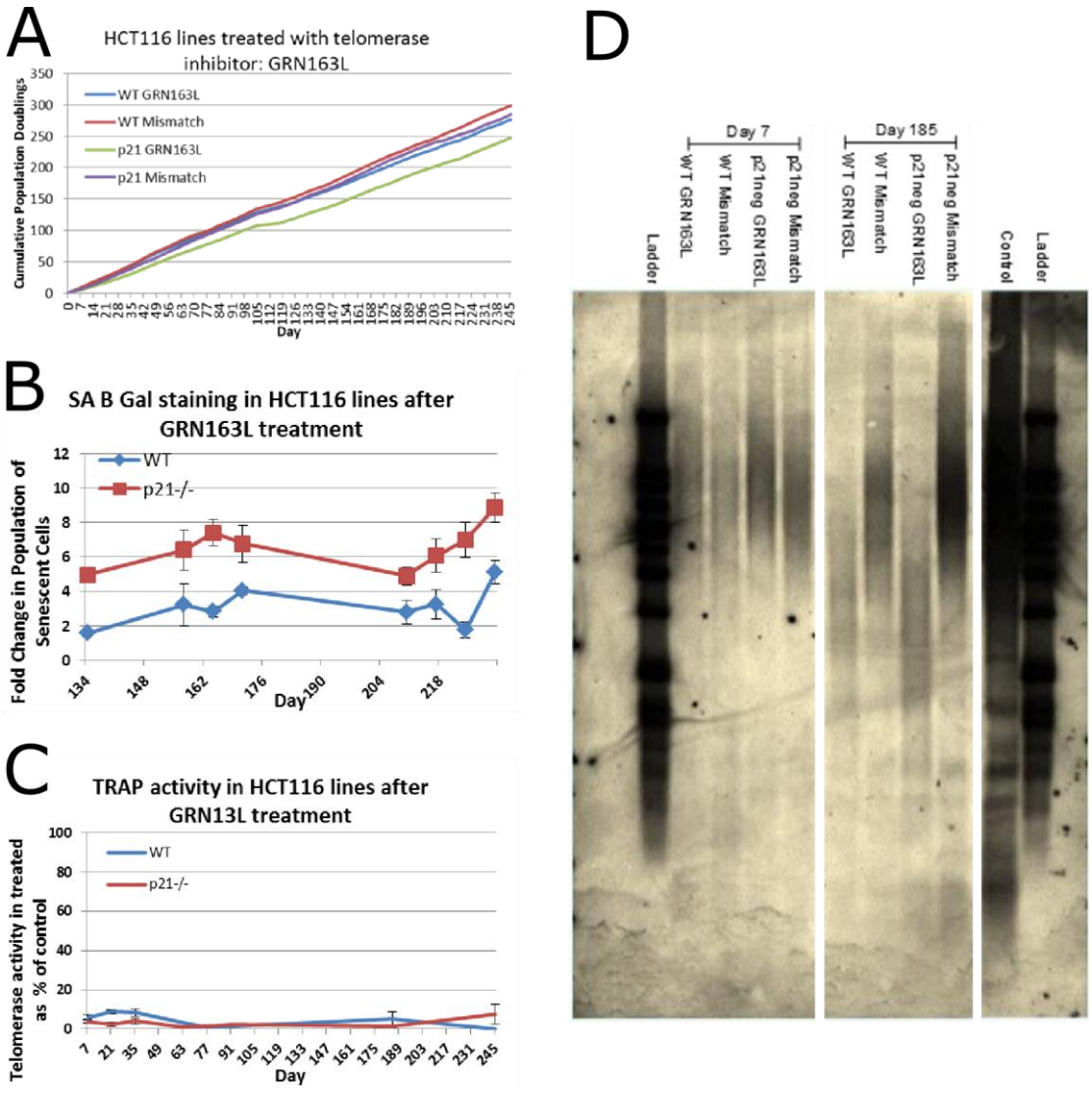


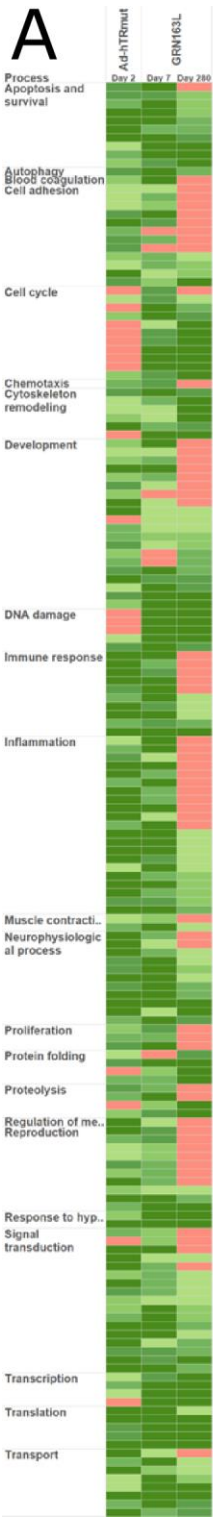
Figure 6.1. Long term treatment of HCT116 WT and HCT116 p21-/- with telomerase inhibitor GRN163L.

(A) Cumulative population doublings of HCT116 cell lines during weekly treatments with 5uM GRN163L over a long time course. (B) Senescent phenotype of HCT116 cell lines after treatment with GRN163L detected by SA B Gal staining. (C) TRAP assay for telomerase activity as a percentage of control over the time course. (D) Telomere restriction fragment southern blot for telomere length at day 7 or day 184 post GRN163L treatment in WT and p21$^{-/-}$ lines where Mismatch is the control.

6.2.2. Specific process network profiles are associated with telomere uncapping and telomerase inhibition

In order to identify possible pathways responsible for the different effects caused by different dysfunction triggers, PNO analysis was performed comparing WT cells after long term telomerase inhibition or telomere uncapping with Ad-hTRmut treatments. Figure 6.2A shows the full process network profile for WT after Day 7 or Day 280 GRN163L treatment and Day 2 of Ad-hTRmut treatment. It can be seen that the profile after day 7 of telomere shortening had very few significantly enriched networks indicating few expression changes by this time point however by day 280 a large number of networks had become significant. By the early time point after Ad-hTRmut treatment there were also a number of significantly enriched networks however these were in different biological processes than those altered after day 280 of GRN163L treatment.

Figure 6.2B shows a filtered subset of the overall process network profile where biological processes which had no significant process networks have been excluded. GRN163L treatment resulted in enrichment of process networks in Cell adhesion, Immune Response, Inflammation, Proliferation and Reproduction. However, Ad-hTRmut treatment resulted in enrichment of Cell cycle and DNA damage process networks. Also, GRN163L treatment required an extended treatment period before changes in expression were seen. However, Ad-hTRmut caused a rapid response with expression changes seen only 2 days post treatment which is indicative of the rapid uncapping nature of the model. Although GRN163L appeared to cause a population of cells in each culture to become senescent, inhibited telomerase and induced telomere shortening the compound failed to reduce the growth rate of the cultures significantly. In addition, under prolonged treatment the resulting profile did not resemble that of Ad-hTRmut treated cells indicating that the compound is not achieving its expected mechanism of action which would be eventual telomere uncapping due to telomere attrition.



B

Process	network	Ad-hTRmut Day 2	GRN163L Day 7	GRN163L Day 280
Cell adhesion	Attractive and repulsive receptors	0.21	0.05	0.00
	Synaptic contact	0.10	0.43	0.00
	Platelet-endothelium-leucocyte interactions	0.13	0.41	0.00
	Leucocyte chemotaxis	0.77	0.01	0.00
	Glycoconjugates	0.95	0.73	0.01
	Platelet aggregation	0.48	0.04	0.01
	Amyloid proteins	0.72	0.46	0.03
	Cell-matrix interactions	0.68	0.03	0.05
	Integrin-mediated cell-matrix adhesion	0.32	0.51	0.17
	Integrin priming	0.15	0.59	0.25
	Cadherins	0.98	0.17	0.27
	Cell junctions	0.79	0.37	0.92
Cell cycle	G1-S Growth factor regulation	0.01	0.79	0.00
	G1-S Interleukin regulation	0.16	0.64	0.10
	Meiosis	0.00	1.00	0.57
	G0-G1	0.25	1.00	0.68
	G2-M	0.00	0.23	0.90
	S phase	0.00	0.70	0.98
	G1-S	0.00	0.73	1.00
	Mitosis	0.00	1.00	1.00
	Core	0.00	1.00	1.00
	Spindle microtubules	0.00	1.00	1.00
	Proteolysis in cell cycle and apoptosis	0.25	0.64	0.96
DNA damage	Checkpoint	0.00	1.00	0.92
	DBS repair	0.05	1.00	0.94
	MMR repair	0.00	1.00	0.98
	BER-NER repair	0.08	1.00	1.00
	Core	0.76	1.00	0.78
Immune response	Phagocytosis	0.99	0.84	0.00
	Th17-derived cytokines	0.99	0.55	0.01
	BCR pathway	0.99	0.67	0.02
	Phagosome in antigen presentation	0.90	1.00	0.02
	Lymphocyte proliferation	0.79	0.82	0.02
	TCR signaling	0.45	1.00	0.10
	T helper cell differentiation	0.97	0.68	0.12
	IL-5 signaling	0.88	1.00	0.15
	Antigen presentation	0.47	0.80	0.52
	Innate immune response to RNA viral infection	0.92	1.00	1.00
Inflammation	Kallikrein-kinin system	0.17	1.00	0.00
	Interferon signaling	0.61	1.00	0.00
	Complement system	0.68	0.11	0.00
	IL-4 signaling	1.00	1.00	0.00
	Amphoterin signaling	0.93	0.61	0.00
	IFN-gamma singaling	0.61	1.00	0.00
	IL-10 anti-inflammatory response	0.93	0.50	0.02
	IL-6 signaling	1.00	0.62	0.02
	IL-2 signaling	0.99	1.00	0.02
	Jak-STAT Pathway	0.90	0.19	0.03
	MIF signaling	0.51	1.00	0.04
	IgE signaling	0.98	1.00	0.06
	Innate inflammatory response	0.98	1.00	0.06
	Histamine signaling	0.94	1.00	0.07
	TREM1 signaling	0.97	0.69	0.10
	Protein C signaling	0.14	1.00	0.15
	Neutrophil activation	0.90	0.52	0.31
	IL-12,15,18 signaling	1.00	1.00	0.33
	IL-13 signaling pathway	0.99	0.52	0.37
	NK cell cytotoxicity	0.80	0.74	0.40
	Inflammasome	1.00	1.00	0.58
Proliferation	Negative regulation of cell proliferation	0.38	0.78	0.00
	Positive regulation of cell proliferation	0.62	0.53	0.01
	Lymphocyte proliferation	0.79	0.82	0.02
Reproduction	GnRH signaling pathway	0.81	0.74	0.00
	Gonadotropin regulation	0.61	0.80	0.00
	FSH-beta signaling pathway	0.08	0.37	0.00
	Male sex differentiation	0.20	0.31	0.00
	Spermatogenesis, motility and copulation	0.66	0.55	0.01
	Progesterone signaling	0.32	0.51	0.05
	ESR1-membrane pathway	0.81	0.52	0.05
	ESR1-nuclear pathway	0.34	0.09	0.09
	ESR2 pathway	0.89	1.00	0.46
	Feeding and Neurohormone signaling	0.56	1.00	1.00

Figure 6.2. Process network profile analysis after either GRN163L or hTRmut treatments in HCT116 WT cells.

(A) Full process network profile of HCT116 WT cells 7 or 280 days post GRN163L treatment and day 2 post Ad-hTRmut treatment where red rectangles are networks where $p < 0.05$. Ad-hTRmut treated samples were the same as in previous figures. (B) Filtered profile view where insignificant networks have been excluded.

6.2.3. Hypoglycaemia sensitises HCT116 WT cells to GRN163L

This work was carried out in collaboration with Dr Alan Bilsland, University of Glasgow. The experiments in Figures 6.3B, C and D were carried out by Dr Alan Bilsland.

After comparing the effects of independent dysfunction triggers and finding the effects of telomerase inhibition to be weak and unable to produce a strong phenotype in the HCT116 cell lines we sought out ways to enhance the effects of telomerase inhibition. Our laboratory has recently demonstrated that hypoglycaemia reduces telomerase activity and induces telomere shortening. Therefore it was hypothesised that the effects of GRN163L could be enhanced by treating cells under glucose restricted conditions. Figure 6.3A shows telomerase activity in HCT116 cells after titration of GRN163L or Mismatch control measured 2 days post treatment under physiological conditions. GRN163L inhibited telomerase in the 1-10 μM range and at concentrations below 1 μM had little to no effect on telomerase activity.

In order to characterise the effects of Hypoglycaemia on telomerase activity HCT116 WT cells were cultured at a range of Glucose concentrations (Figure 6.3B). It was found that culturing HCT116 WT cells in media containing 1 mM Glucose or less, reduced telomerase activity by up to 50%. To assess the effect of GRN16L under hypoglycaemic conditions a long term culture was set up in Glucose free media using 200 nM GRN163L (Figure 6.3C). 200 nM was chosen as telomerase was not inhibited at this concentration under physiological glucose conditions. Under hypoglycaemic conditions, GRN163L strongly inhibited telomerase activity when compared to hypoglycaemic control suggesting that hypoglycaemia sensitises the cells to telomerase inhibition. In addition, hypoxia did not significantly reduce the growth rate of HCT116 WT cells (Data not shown).

To further characterise the extent of this sensitisation, telomere length was measured after treatment under hypoglycaemic conditions. In Figure 6.3D it can be seen that by day 77, hypoglycemia had caused a mild shortening of the

telomeres. In samples treated with 200 nM GRN163L under hypoglycaemic conditions this shortening effect was enhanced indicating that hypoglycaemia enhances the telomere shortening effect of GRN163L.

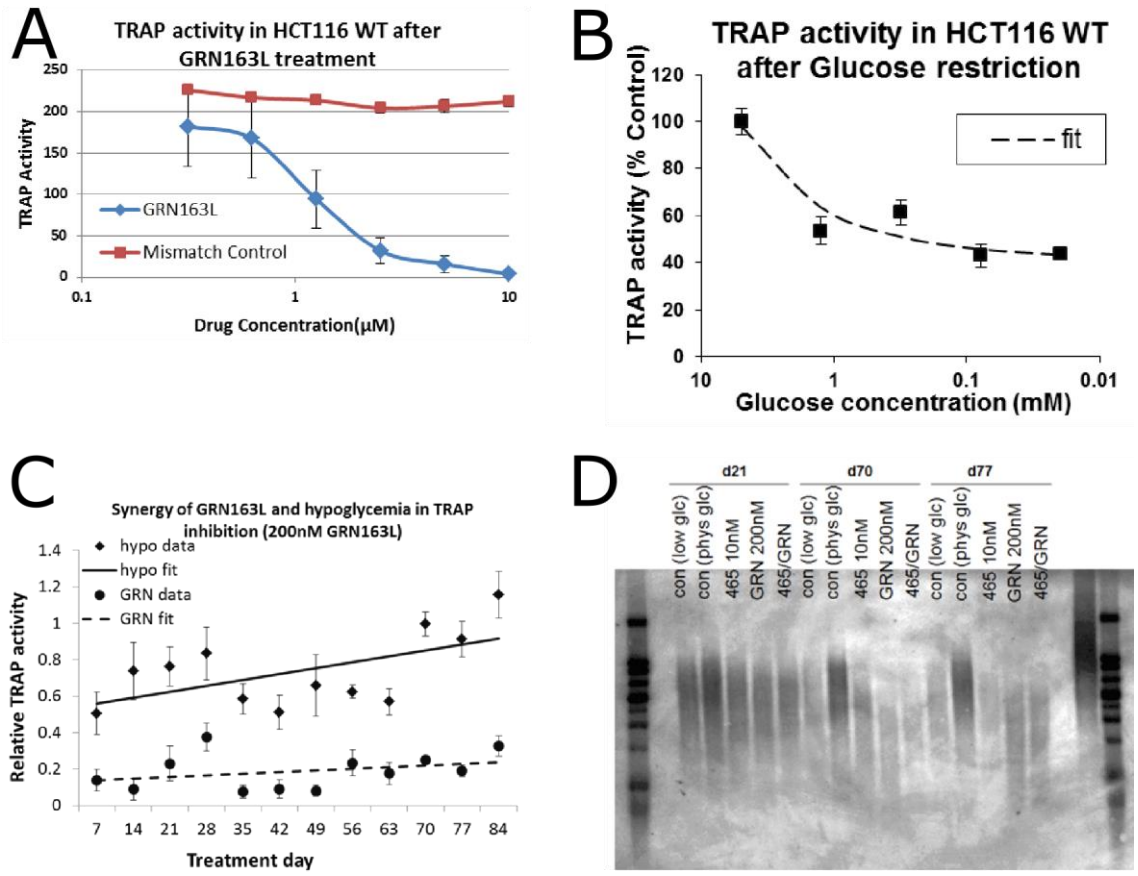


Figure 6.3. Telomerase activity and telomere length after GRN163L treatment in HCT116 cell lines under physiological or hypoglycaemic conditions.

(A) TRAP assay after titration of GRN163L in HCT116 cells 2 days post treatment under physiological conditions. (B) TRAP assay for telomerase activity as a percentage of control over the time course. (C) TRAP assay for telomerase activity as a percentage of control over the time course in HCT116 WT after 200nM GRN163L treatments under hypoglycaemic conditions. (D) Telomere restriction fragment southern blot for telomere length at day 21, 70 and 77 post GRN163L or control treatments in HCT116 WT cell line.

To further characterise the processes affected after telomerase inhibition under hypoglycaemic conditions PNO analysis was performed. HCT116 WT cells treated with GRN163L under physiological conditions were compared with those treated under hypoglycaemic conditions and also with HCT116 WT cells treated with Ad-hTRmut to uncap the telomeres. Figure 6.4A shows the full process network profile while Figure 6.4B shows a filtered version of the profile. Under both physiological and hypoglycaemic conditions there were few significant networks by day 7, post treatment. However, under physiological conditions it wasn't until day 280 that many significant processes emerged, whilst under hypoglycaemic conditions there were many significant networks by day 84. Interestingly, there was little overlap in the profiles between physiological GRN163L samples and hypoglycaemic GRN163L samples. However, the hypoglycaemic GRN163L samples overlapped significantly with the hTRmut treated samples. This visual interpretation was confirmed by hypergeometric analysis as shown in Figure 6.4C. These results suggest that hypoglycaemia causes a shift in molecular response to telomerase inhibition, resulting in activation of similar pathways to those involved in rapid telomere uncapping, specifically DNA damage and Cell Cycle pathways.

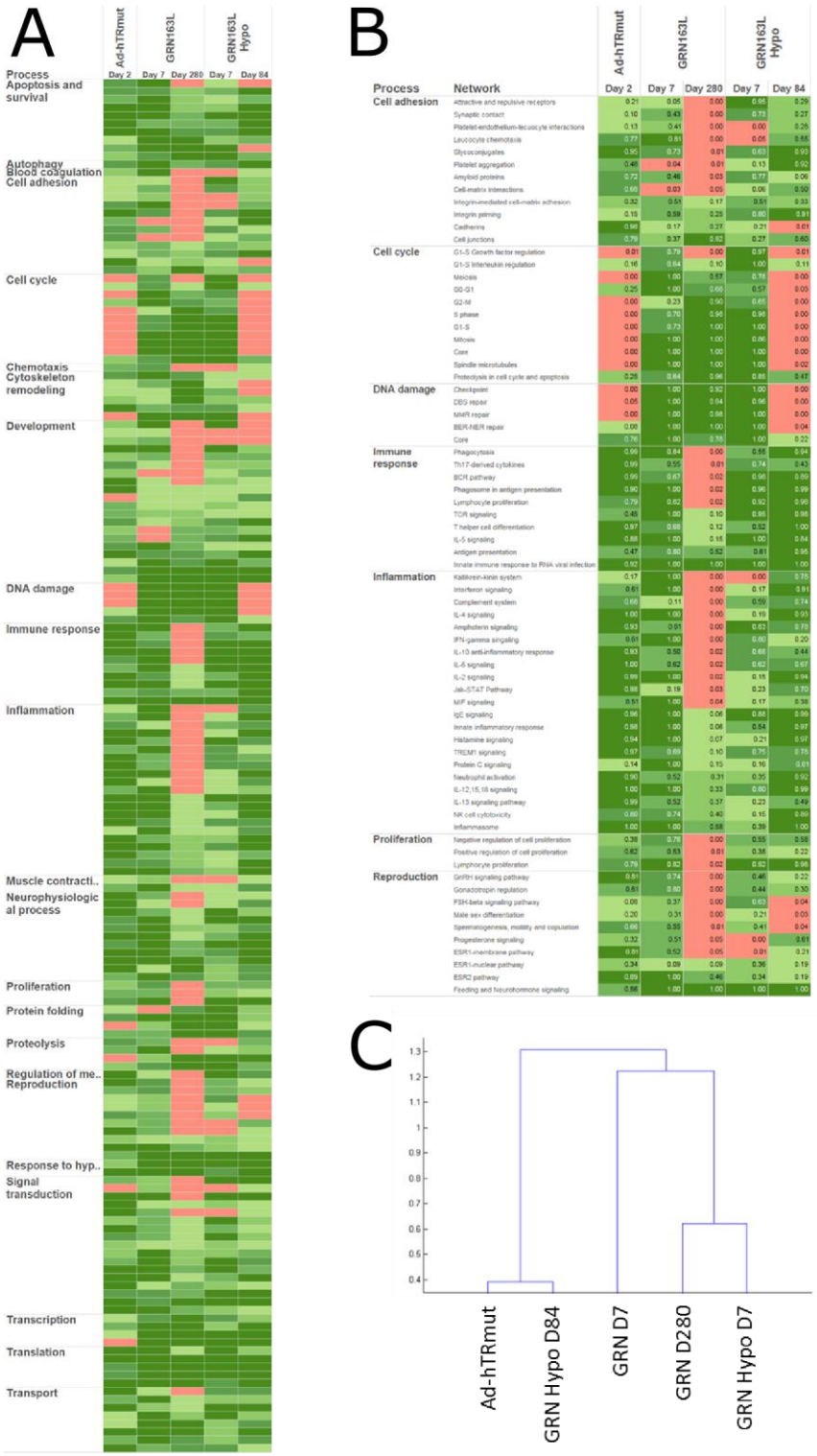


Figure 6.4. Process network analysis comparing different telomere dysfunction triggers.

(A) Full process network profile of HCT116 WT cells 7 or 280 days post GRN163L treatment, 7 or 84 days post GRN163L + hypo treatment and day 2 post hTRmut treatment where red rectangles are networks where p <0.05 (B) Filtered profile view where insignificant networks have been excluded. (C) Hypergeometric analysis of the process network profile.

6.3. Discussion

6.3.1. HCT116 cell lines are sensitive to telomere uncapping but resistant to telomerase inhibition

Induction of rapid telomere uncapping by Ad-hTRmut had a clear difference in rate of growth reduction in the WT, $p21^{-/-}$ and $p53^{-/-}$ backgrounds as shown in Chapter 4. In contrast, telomerase inhibition by GRN163L only had a mildly stronger effect on growth of the $p21^{-/-}$ cells relative to WT over a long time course. Further to this, analysis of the senescence phenotype after treatment with both triggers highlighted that after rapid induction of telomere dysfunction $p21^{-/-}$ cells had less SA ß Gal staining relative to WT cells. However, after long term telomerase inhibition the SA ß Gal staining was greater in $p21^{-/-}$ cells. Despite this apparent increase in senescent $p21^{-/-}$ cells the cultures continued to grow past day 250 of culture suggesting that a significant population of cells were resistant to the effects of telomerase inhibition. Also, although cell adhesion pathways were implicated after GRN163L treatment and anti-adhesive effects have been shown in the past this effect was not observed in any of the cultures (Jackson et al., 2007). Cells were given appropriate time to adhere to the plates before treatment and throughout the time course adhesion to the culture flasks did not appear to be an issue or cause the differences in population doublings observed.

Despite the inability of telomerase inhibition by GRN163L to cause a reduction in growth, a number of pathways did show significant expression changes over the time course. Outcomes of telomere shortening may be due to changes in telomere state rather than reaching critical length (Karlseder et al., 2002). Therefore, by simply disrupting telomere regulation by telomerase inhibition it may be possible to activate anti proliferative pathways. However, this was not observed in the HCT116 cell lines as they continued to grow for a sustained period of time even during ablation of telomerase activity. This suggests that changes in expression are likely to be compensating for telomerase inhibition rather than initiating processes which would lead to senescence or apoptosis. Therefore, this would implicate pathways enriched in response to GRN163L treatment across the time course such as Inflammation and Immune response

process networks. While this may be another route to resisting telomere dysfunction, this profile may also be indicative of off target effects or early indicators toxicity which may be a result of using relatively high concentrations of GRN163L (5uM). Therefore a sensitisation method was desired which may reduce the concentration of GRN163L required to inhibit telomerase activity to a sub-toxic level.

The drug and mutant have different outcomes, likely due to their mechanism of action. AdhTRmuts rapid response observed in the lines is likely due to rapid induction of DNA damage signalling pathways as a result of uncapping of the telomeres. This was reflected in the pathways which became enriched in response such as Cell cycle and DNA damage associated networks, whereas GRN163L causes shortening at a rate relative to the proliferation rate of the cells. Due to this, not all cells are at equal telomere length at all times. Therefore, not all cells will have an identical response which was represented in the continued proliferation of the cells in culture despite continued GRN163L treatment. It was noted that GRN163L caused enrichment of inflammation and immune response pathways. Also, there was a population of senescent cells observed. This could indicate that SASP pathways are being activated and this enrichment is reflective of senescence associated secretory phenotype. However, it did not appear to be enough to cause collapse of the entire culture, and cells continued to proliferate. However, the drug does lead to a senescent population of cells over time, indicating that the telomeres are reaching critical length. ChIP was performed for TRF2 after induction of telomere uncapping by AdhTRmut. It may have been useful to perform this experiment after treatment with GRN163L to assess whether shelterin was being disrupted, which would be the case if the telomeres were reaching critical length. Further to this, if shelterin was disrupted one would expect to see induction of DNA damage signalling responses. In the WT line there was enrichment in DNA damage response pathways in response to telomere uncapping by AdhTRmut but not by GRN163L treatment. However, treatment under hypoglyceaemic conditions did result in enrichment in DNA damage response pathways. This warrants further investigation and assays to look for DNA damage foci and chromosomal fusions would be suitable steps to take to further characterise the effects of telomere

uncapping and telomere inhibition under physiological and hypoglycaemic conditions.

6.3.2. Hypoglycaemia sensitises HCT116 cells to telomerase inhibition and alters regulatory pathways

Our lab identified hypoglycaemia as an environmental telomere dysfunction trigger which reduces telomerase activity and induces telomere shortening. In addition, glucose restriction has been shown to sensitise breast cancer cells to telomerase inhibition by another compound whilst glucose substitution in telomere dysfunction mice results in increased lifespan (Missios et al., 2014; Wardi et al., 2014). Indeed we found that glucose restriction was able to sensitise a telomerase inhibition resistant cell line, HCT116 WT, by reducing the drug concentration required for effective telomerase inhibition. This also appeared to enhance the rate of telomere shortening. Further investigation revealed that a hypoglycaemic environment caused a shift in expression profile in HCT116 WT cells when treated with GRN163L resembling telomere uncapping by Ad-hTRmut. Part of this shift could have been due to reduced toxicity at a lower concentration of GRN163L as inflammatory and immune response pathways were no longer being enriched. However, the shift towards enrichment of DNA damage and Cell Cycle pathways was not accompanied by any reduction in growth rate of the cells and would require further study. One possibility is that normal glucose metabolism is required for maintenance of the telomeres and sustaining telomere integrity. Therefore, under telomerase inhibition and under glucose restriction there is a shift in balance towards cellular senescence. This could indicate that glucose restriction causes changes in the regulatory pathways or signalling responses to telomere dysfunction by telomere shortening.

It may also be possible that the HCT116 line is resistant to telomerase inhibition as other cell lines have demonstrated significantly reduced lifespans under GRN163L treatment (Burchett et al., 2014). Nonetheless, being able to sensitise previously unresponsive lines is a desirable outcome for further development of the compound as a therapeutic. This could be achieved through co-treatment with existing drugs such as Metformin, used in treatment of type 2 diabetes,

which suppresses glucose production by the liver. An alternative method could be to inhibit glucose metabolism in another way. This has been investigated previously but not from the angle of telomerase inhibition as a co-treatment option (Munoz-Pinedo et al., 2003).

Both hypo and hyperglycaemia have been implicated in regulation of both telomeres and telomerase. Hypoglycaemia has been associated with reduced telomerase activity (Wardi et al., 2014). In addition, a recent in vitro study has shown that inhibition of hTERT expression in muscle cells cause a reduction in basal glucose uptake, whereas overexpression of TERT significantly increased their glucose uptake. This effect of hTERT occurs outside the nucleus through direct interaction of hTERT with glucose transporters and therefore Telomerase and telomere regulation appears to be directly linked to glucose metabolism (Shaheen et al., 2014). It also appears that the mitochondrial electron transport chain function is altered during hypoglycaemia due to the decreased availability of reducing equivalents resulting in increased production of reactive oxygen species (ROS) by mitochondria under hypoglycaemic conditions (McGowan et al., 2006). Furthermore, reducing glucose may then push cells into alternative metabolic routes, such as lactate metabolism which produces more ROS. Overall, hypoglycaemia reduces telomerase activity whilst increasing oxidative stress on the telomeres and enhances telomere shortening.

Hyperglycaemia increases ROS from the mitochondrial electron transport chain. In addition, increased glucose auto-oxidation, activation of the polyol pathway and protein kinase C pathway, and production of advanced glycation end products also play roles in increasing the level of oxidative stress (Araki and Nishikawa, 2010). Such exacerbated oxidative stress accelerates the shortening of telomeres, (Jennings et al., 2000). Mouse models which are maintained on a high caloric intake, incorporating glucose, demonstrate increased ROS levels in adipose tissues (Minamino et al., 2009). Furthermore, it has been observed in humans that type 2 diabetes mellitus (T2DM) patients show a reduction of ß-cell mass in the pancreas (Sakuraba et al., 2002). Moreover, in a mouse model maintained on a diabetogenic diet it was demonstrated that ß-cells significantly declined after 12 months and that this was due to accelerated cellular

senescence and apoptosis (Sone and Kagawa, 2005). It has also been demonstrated that glycaemic control reduced telomere shortening in the leukocytes of patients with both T1DM and T2DM (Uziel et al., 2007). This was corroborated in another similar study in patients with T2DM who has used agents to maintain their glucose levels demonstrated longer telomeres in ß-cells than those who did not (Tamura et al., 2014). Therefore hyperglycaemia also appears to increase telomere shortening through oxidative stress. Overall, metabolism is a growing field of research in cancer and therapeutic options. From the perspective of telomeres however, the role of metabolism is not well understood. Telomeres are sensitive to oxidative stress, and both hyper and hypoglycaemia appears to increase ROS levels within cells, which may be the mechanism by which impaired glucose metabolism enhances sensitivity to telomere dysfunction.

6.3.3. Clinical opportunities with GRN163L

Telomerase expression is a common trait in malignant cancers and therefore is an enticing therapeutic target. However, specific ablation of telomerase has not proven as effective as expected. All of the HCT116 isogenic lines continued to proliferate beyond 280 days of continued GRN163L treatment. Sensitisation was achieved through treatment under hypoglycaemic conditions but not to the degree that proliferation was halted entirely. However, this is still encouraging and demonstrates that further investigation is required with respect to telomerase inhibition and a number of challenges remain to be solved to further develop telomerase inhibition as a therapeutic strategy. One issue with GRN163L is with the half-life of the drug in human tissues. The tissue half-life of GRN163L, or the time it takes for the concentration or amount of GRN163L to be reduced by half, in bone marrow, spleen, liver and tumour has been estimated to be 41 hours in humans, based on data from animal studies and clinical trials. By day 280 of treatment a number of inflammation and immune response pathways demonstrated enrichment and therefore, toxicity and chronic inflammation could be a concern given the high dose required to ablate telomerase activity in human cells. Therefore, sensitisation options such as a hypoglycaemic environment may be desirable in order to reduce these effects

while maintaining a desired level of telomerase inhibition. Indeed, hypoglycaemia did significantly reduce the required dose of GRN163L to reduce telomerase activity to below 10% of the activity in untreated cells. Therefore one approach in the clinic may be to co-treat with Metformin, which reduces glucose production in the liver, or to treat under dietary conditions in which glucose intake is restricted. Other co-treatment options may also be of value. It has been demonstrated that telomerase inhibition can enhance chemosensitivity of tumours (Wardi et al., 2014). Therefore from a clinical perspective, telomerase inhibition may be worth further investigation, as a co-treatment option with current therapeutics.

Another challenge is in the timing of GRN163L treatment. Due to the latent effect of the drug, ideally, treatments would occur as early as possible in order for the telomeres to shorten to the point where proliferation arrest occurs. One thought is that combining chemotherapy with telomerase therapy would be predicted to both shorten telomeres and reduce tumour burden. Even if rare cancer stem cells are quiescent, eventually they will have to proliferate to maintain the growth of the tumour. The presence of a non-toxic or minimally toxic dose of a telomerase inhibitor should then affect the telomerase-positive cancer stem cells and eventually lead to apoptosis. Another issue is that GRN163L relies on cells replicating their DNA in order for telomeres to shorten in the absence of telomerase activity, and therefore the tumours must be proliferative. However, at a late stage, allowing tumours to continue to proliferate could put patients at risk. Therefore, early stage treatment would be the ideal time to treat with a telomerase inhibitor while the tumour mass is not yet at a high risk to the patient.

Chapter 7: Further overall Discussion

7.1. Telomere biology and pathways of interest

As discussed previously, during development of the PNO analysis method I was able to apply the method to data sets supplied by Dr Bilsland and Dr Degerman and take analysis of their data sets beyond traditional analysis approaches. In particular, where traditional approaches would have required further analysis in order to identify pathways of interest my novel approach simplified the process and enabled identification of pathways as a standard data output. After validating the viability of the methodology in these two studies I was then able to use the method to analyse my own data sets and explore telomere biology further. Prior to array analysis, it was identified that isogenic background led to differences in phenotype and expression profile to two different telomere dysfunction triggers. It was identified that p21 was central to sensitivity to telomere dysfunction, and that loss of p21 sensitised cells to both telomere uncapping and telomere shortening via telomerase inhibition, demonstrating reduced cell growth and increased senescent cell population. Through PNO analysis it was identified that $p21^{-/-}$ cells demonstrated a reduction in regulation of DNA damage response pathways. Further to this, $p21^{-/-}$ cells were unable to down regulate spindle and centromere formation genes which may suggest cell death through another mechanism, such as mitotic catastrophe. During this analysis mismatch repair was also identified as a network of interest and further study.

Analysis of the mismatch repair pathway revealed that $p21^{-/-}$ cells were also unable to engage this pathway. Western blotting revealed that WT cells maintained MutSα levels whereas this was reduced in the $p21^{-/-}$ cell line after induction of telomere uncapping. In addition, siRNA mediated knockdown of, MSH2, a component of MutSα, caused the phenotype of the WT to shift and resemble the $p21^{-/-}$ cell line. This was demonstrated both in the senescent phenotypic response to telomere uncapping but also in the shift in expression profile and enrichment of networks observed in the associated PNO analysis. Therefore it was identified that p21 appears to be central to regulation of

telomere dysfunction through a regulatory mechanism involving the mismatch repair pathway.

These findings presented a number of opportunities. Firstly, they demonstrated the strength in the methodology used and the ability to identify pathways for further research. Secondly, they presented mismatch repair and association of the process with telomere dysfunction as an area for further study. The work performed did however have a number of weaknesses. One of the major weaknesses was in the range of phenotypes analysed. Growth rate and senescence were evaluated in response to telomere dysfunction. However, an obvious next step would have been to evaluate apoptosis after induction of telomere uncapping in the isogenic cell lines which was not performed. In addition, it would have been valuable to evaluate the stage at which the different cell lines stalled in the cell cycle, as this may have informed the route to cell death as mitotic catastrophe was hypothesised as a possible route in $p21^{-/-}$ cell lines. It would have also been of value to look at the DNA damage response and to evaluate the predominant forms of DNA damage in the isogenic cell lines. For example, it would have been useful to look at DNA damage foci, accumulation of γH2AX and for chromosomal fusions. This would have informed as to the DNA damage status of the cells, as it was postulated that $p21^{-/-}$ may have been deficient in maintenance of DNA damage leading to cell death. The Shelterin status of the cells was also not fully established. It was determined that TRF2 was reduced at telomeres after induction of telomere dysfunction, however, the other Shelterin components were not evaluated. With respect to mismatch repair, the exact interactions were also not evaluated. It would have been useful to identify where MSH2 was localising within the cells. For example, MSH2 may have played a direct role at telomeres or may have had a secondary signalling role.

7.2. Clinical relevance of targeting telomeres

Telomeres are an interesting clinical target in cancer treatment due to activation of telomerase being a common method of bypassing the immortalisation barrier by continued maintenance of telomere length.

Therefore, evaluation of the telomerase inhibitor GRN163L was of interest. Comparison against telomere uncapping by Ad-hTRmut revealed both differences in senescent and growth phenotypic response and expression profile to different telomere dysfunction triggers, indicating that the method of inducing telomere dysfunction has varied biological consequences. However, of interest was the ability to sensitise cells to telomere dysfunction through introduction of the environmental trigger of hypoglycaemia. Telomerase therapies have had varied results in the clinic. GRN163L in particular has demonstrated a wide range of efficacies in different cell lines. In our hands, GRN163L was not effective in reducing the growth rate of HCT116 cell lines independently. However, it was revealed that hypoglycaemia enabled sensitisation of HCT116 cell lines to telomerase inhibition. This was exciting as it may point towards co-treatments or treatments with existing cancer treatment modalities as a path forward for targeting telomerase during cancer treatment.

However, the work done here is merely a starting point. Further work is required to determine the effectiveness of co-treatments with telomerase inhibition. In the work done here further phenotypic analysis would have been beneficial. For example, the apoptosis status of the cell lines was not evaluated. This may have helped differentiate the effects of telomerase inhibition under physiological conditions with telomerase inhibition under hypoglycaemic conditions. In addition, introduction of hypoglycaemia was only evaluated through glucose restriction in cell culture through changed media conditions, i.e. use of glucose free media. However, a number of glucose metabolism mediating compounds are available, such as metformin. It could have been beneficial to evaluate other methods of glucose restriction which may have been clinically applicable to further establish the effectiveness of glucose restriction on telomerase activity.

7.3. Conclusion

Telomere dysfunction is complex and involves interaction between a wide range of biological processes. Bypass of telomere dysfunction is essential in oncogenesis and the immortalisation process and enhancing our understanding of telomere dysfunction is essential (Prieur and Peeper, 2008). Due to the range of signalling interactions involved a systems biology approach is required in order

to characterise the biological pathways involved in telomere dysfunction. However, systems biology itself is a continually developing field with its own set of challenges. Management, visualisation and interpretation of large data sets is still a difficult problem and although the field has made great strides there is still room for improvement (O'Donoghue et al., 2010a).

During this project I developed a novel methodology and workflow for analysis of data sets comprised of gene lists. Most traditional approaches focus on identification of small sets of genes that may or may not be in a pathway. However, the advantage of my novel approach is that it allows for identification of biological networks leading to immediate identification of pathways and direct characterisation of new biology. The effectiveness of this approach was demonstrated in two separate collaborations. In the first collaboration with Dr Alan Bilsland I was able to assist in the identification of putative mechanisms of action for a novel senescence inducing compound. This was achieved by comparing expression data across a panel of kinase inhibitors with which the compound was suggested to share a mechanism of action. The result was confirmation that the compound operated through an alternative mechanism to the other compounds in the panel. Furthermore, the results helped to direct investigation of the mechanism of action. The flexibility of the approach was then highlighted in the second collaboration with Dr Sofie Degerman and her group at Umeå University. During this collaboration an alternative data set comparing differentially methylated genes in a range of T cell lines over a time course (Degerman et al., 2014), the analysis I performed assisted in identification of biology associated with the immortalisation process and the role played by methylation.

Further to use of the methodology to assist in collaborative efforts was its application to data sets within our own laboratory group. The primary aim of this study was to characterise telomere dysfunction regulatory pathways and identify previously unknown biology in the process. In the first instance this was achieved through use of the methodology to analyse the effect of genetic background on regulation of telomere dysfunction. Comparison of HCT116 WT, $p21^{-/-}$ and $p53^{-/-}$ cell lines response to telomere uncapping by Ad-hTRmut revealed p21 to be

essential in response to telomere uncapping. $p21^{-/-}$ cells demonstrated enhanced sensitivity to telomere uncapping and through the PNO analysis method Mismatch Repair was identified as a candidate process which may be involved in regulation of telomere dysfunction. Further validation of the mismatch repair pathway indeed confirmed it to have a role in managing sensitivity to telomere dysfunction and p21 appeared to be essential in induction of components of mismatch repair pathway in response to telomere uncapping.

In addition to investigating the role of genetic background in telomere dysfunction the method was used to investigate the effects of the telomerase inhibitor GRN163L and the synergistic effects of telomerase inhibition during hypoglycaemia. GRN163L on its own was unable to reduce the growth rate of HCT116 cell lines over a long treatment period. Long term treatment did result in enrichment in a number of process networks. However, the processes affected were not traditionally associated with outcomes of telomere dysfunction indicating that the compound was not achieving its expected mechanism of action. Hypoglycaemia was found by our group to reduce telomerase activity and cause telomere shortening. It was also demonstrated to enhance the effects of a telomerase inhibitor by another group (Wardi et al., 2014). Therefore, we decided to investigate the synergistic effects of treating GRN163L cells under hypoglycaemic conditions. Indeed, we found that hypoglycaemia significantly reduced the required concentration of the compound for effective telomerase inhibition and mildly increased the rate of telomere shortening over a small time scale. So while recent telomerase inhibitors have not been successful in the clinic, the options available for telomerase targeted therapy are not yet exhausted.

Overall I feel that this new methodology is an excellent method of complementing existing methods of analysing gene lists and has contributed to the identification of new biology in a range of projects.

References

Abbas, T., and Dutta, A. (2009). p21 in cancer: intricate networks and multiple activities. Nat Rev Cancer *9*, 400-414.

Acosta, J.C., O'Loghlen, A., Banito, A., Guijarro, M.V., Augert, A., Raguz, S., Fumagalli, M., Da Costa, M., Brown, C., Popov, N.*, et al.* (2008). Chemokine signaling via the CXCR2 receptor reinforces senescence. Cell *133*, 1006-1018.

Aliouat-Denis, C.M., Dendouga, N., Van den Wyngaert, I., Goehlmann, H., Steller, U., van de Weyer, I., Van Slycken, N., Andries, L., Kass, S., Luyten, W*., et al.* (2005). p53-independent regulation of p21Waf1/Cip1 expression and senescence by Chk2. Molecular cancer research : MCR *3*, 627-634.

Anderson, B.H., Kasher, P.R., Mayer, J., Szynkiewicz, M., Jenkinson, E.M., Bhaskar, S.S., Urquhart, J.E., Daly, S.B., Dickerson, J.E., O'Sullivan, J*., et al.* (2012). Mutations in CTC1, encoding conserved telomere maintenance component 1, cause Coats plus. Nature genetics *44*, 338-342.

Araki, E., and Nishikawa, T. (2010). Oxidative stress: A cause and therapeutic target of diabetic complications. J Diabetes Investig *1*, 90-96.

Armanios, M., and Blackburn, E.H. (2012). The telomere syndromes. Nature reviews Genetics *13*, 693-704.

Armanios, M., Chen, J.L., Chang, Y.P., Brodsky, R.A., Hawkins, A., Griffin, C.A., Eshleman, J.R., Cohen, A.R., Chakravarti, A., Hamosh, A*., et al.* (2005). Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. Proceedings of the National Academy of Sciences of the United States of America *102*, 15960-15964.

Ballew, B.J., Joseph, V., De, S., Sarek, G., Vannier, J.B., Stracker, T., Schrader, K.A., Small, T.N., O'Reilly, R., Manschreck, C*., et al.* (2013). A recessive founder mutation in regulator of telomere elongation helicase 1, RTEL1, underlies severe immunodeficiency and features of Hoyeraal Hreidarsson syndrome. PLoS genetics *9*, e1003695.

Bar-Yam, Y., Harmon, D., and de Bivort, B. (2009). Systems biology. Attractors and democratic dynamics. Science *323*, 1016-1017.

Barrett, T., and Edgar, R. (2006a). Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. Methods in enzymology *411*, 352-369.

Barrett, T., and Edgar, R. (2006b). Mining microarray data at NCBI's Gene Expression Omnibus (GEO)*. Methods Mol Biol *338*, 175-190.

Beier, F., Foronda, M., Martinez, P., and Blasco, M.A. (2012). Conditional TRF1 knockout in the hematopoietic compartment leads to bone marrow failure and recapitulates clinical features of dyskeratosis congenita. Blood *120*, 2990-3000.

Bilsland, A.E., Hoare, S., Stevenson, K., Plumb, J., Gomez-Roman, N., Cairney, C., Burns, S., Lafferty-Whyte, K., Roffey, J., Hammonds, T*., et al.* (2009). Dynamic telomerase gene suppression via network effects of GSK3 inhibition. PLoS ONE *4*, e6459.

Bilsland, A.E., Pugliese, A., Liu, Y., Revie, J., Burns, S., McCormick, C., Cairney, C.J., Bower, J., Drysdale, M., Narita, M*., et al.* (2015). Identification of a Selective G1-Phase Benzimidazolone

Inhibitor by a Senescence-Targeted Virtual Screen Using Artificial Neural Networks. Neoplasia *17*, 704-715.

Blackburn, E.H. (2001). Switching and signaling at the telomere. Cell *106*, 661-673.

Blackburn, E.H., and Collins, K. (2011). Telomerase: an RNP enzyme synthesizes DNA. Cold Spring Harbor perspectives in biology *3*.

Borah, S., Xi, L., Zaug, A.J., Powell, N.M., Dancik, G.M., Cohen, S.B., Costello, J.C., Theodorescu, D., and Cech, T.R. (2015). Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. Science *347*, 1006-1010.

Bracken, A.P., Kleine-Kohlbrecher, D., Dietrich, N., Pasini, D., Gargiulo, G., Beekman, C., Theilgaard-Monch, K., Minucci, S., Porse, B.T., Marine, J.C.*, et al.* (2007). The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. Genes & development *21*, 525-530.

Bradshaw, P.S., Stavropoulos, D.J., and Meyn, M.S. (2005). Human telomeric protein TRF2 associates with genomic double-strand breaks as an early response to DNA damage. Nature genetics *37*, 193-197.

Brault, M.E., and Autexier, C. (2011). ALTered telomeres in response to telomere dysfunction. Cell Cycle *10*.

Burchett, K.M., Yan, Y., and Ouellette, M.M. (2014). Telomerase inhibitor Imetelstat (GRN163L) limits the lifespan of human pancreatic cancer cells. PLoS ONE *9*, e85155.

Burgart, L.J. (2005). Testing for defective DNA mismatch repair in colorectal carcinoma: a practical guide. Archives of pathology & laboratory medicine *129*, 1385-1389.

Buttitta, L.A., and Edgar, B.A. (2007). Mechanisms controlling cell cycle exit upon terminal differentiation. Curr Opin Cell Biol *19*, 697-704.

Cairney, C.J., Bilsland, A.E., Evans, T.R., Roffey, J., Bennett, D.C., Narita, M., Torrance, C.J., and Keith, W.N. (2012). Cancer cell senescence: a new frontier in drug development. Drug Discov Today *17*, 269-276.

Campisi, J. (2013). Aging, cellular senescence, and cancer. Annu Rev Physiol *75*, 685-705.

Campisi, J., and d'Adda di Fagagna, F. (2007). Cellular senescence: when bad things happen to good cells. Nature reviews Molecular cell biology *8*, 729-740.

Capparelli, C., Guido, C., Whitaker-Menezes, D., Bonuccelli, G., Balliet, R., Pestell, T.G., Goldberg, A.F., Pestell, R.G., Howell, A., Sneddon, S.*, et al.* (2012). Autophagy and senescence in cancer-associated fibroblasts metabolically supports tumor growth and metastasis via glycolysis and ketone production. Cell Cycle *11*, 2285-2302.

Carneiro, T., Khair, L., Reis, C.C., Borges, V., Moser, B.A., Nakamura, T.M., and Ferreira, M.G. (2010). Telomeres avoid end detection by severing the checkpoint signal transduction pathway. Nature *467*, 228-232.

Carrieri, G., Marzi, E., Olivieri, F., Marchegiani, F., Cavallone, L., Cardelli, M., Giovagnetti, S., Stecconi, R., Molendini, C., Trapassi, C.*, et al.* (2004). The G/C915 polymorphism of transforming

growth factor beta1 is associated with human longevity: a study in Italian centenarians. Aging Cell *3*, 443-448.

Cech, T.R. (2004). Beginning to understand the end of the chromosome. Cell *116*, 273-279.

Celli, G.B., Denchi, E.L., and de Lange, T. (2006). Ku70 stimulates fusion of dysfunctional telomeres yet protects chromosome ends from homologous recombination. Nature cell biology *8*, 885-890.

Chen, L.Y., and Lingner, J. (2013). CST for the grand finale of telomere replication. Nucleus *4*, 277-282.

Chen, L.Y., Liu, D., and Songyang, Z. (2007). Telomere maintenance through spatial control of telomeric proteins. Mol Cell Biol *27*, 5898-5909.

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

Chen, L.Y., Zhang, Y., Zhang, Q., Li, H., Luo, Z., Fang, H., Kim, S.H., Qin, L., Yotnda, P., Xu, J.*, et al.* (2012b). Mitochondrial localization of telomeric protein TIN2 links telomere regulation to metabolic control. Molecular cell *47*, 839-850.

Chen, Y., Yang, Y., van Overbeek, M., Donigian, J.R., Baciu, P., de Lange, T., and Lei, M. (2008). A shared docking motif in TRF1 and TRF2 used for differential recruitment of telomeric proteins. Science *319*, 1092-1096.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., Dickins, R.A., Narita, M., Zhang, M., and Lowe, S.W. (2010). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. Cancer Cell *17*, 376-387.

Chicas, A., Wang, X., Zhang, C., McCurrach, M., Zhao, Z., Mert, O., Dickins, R.A., Narita, M., Zhang, M., and Lowe, S.W. (2012). Dissecting the unique role of the retinoblastoma tumor suppressor during cellular senescence. Cancer Cell *17*, 376-387.

Chien, Y., Scuoppo, C., Wang, X., Fang, X., Balgley, B., Bolden, J.E., Premsrirut, P., Luo, W., Chicas, A., Lee, C.S.*, et al.* (2011). Control of the senescence-associated secretory phenotype by NF-kappaB promotes senescence and enhances chemosensitivity. Genes & development *25*, 2125-2136.

Chin, K., de Solorzano, C.O., Knowles, D., Jones, A., Chou, W., Rodriguez, E.G., Kuo, W.L., Ljung, B.M., Chew, K., Myambo, K.*, et al.* (2004). In situ analyses of genome instability in breast cancer. Nature genetics *36*, 984-988.

Collado, M., Blasco, M.A., and Serrano, M. (2007). Cellular senescence in cancer and aging. Cell *130*, 223-233.

Collins, K., and Mitchell, J.R. (2002). Telomerase in the human organism. Oncogene *21*, 564-579.

Conde-Perezprina, J.C., Leon-Galvan, M.A., and Konigsberg, M. (2012). DNA mismatch repair system: repercussions in cellular homeostasis and relationship with aging. Oxidative medicine and cellular longevity *2012*, 728430.

Coppe, J.P., Desprez, P.Y., Krtolica, A., and Campisi, J. (2010). The senescence-associated secretory phenotype: the dark side of tumor suppression. Annu Rev Pathol *5*, 99-118.

Coppe, J.P., Patil, C.K., Rodier, F., Sun, Y., Munoz, D.P., Goldstein, J., Nelson, P.S., Desprez, P.Y., and Campisi, J. (2008). Senescence-associated secretory phenotypes reveal cell-nonautonomous functions of oncogenic RAS and the p53 tumor suppressor. PLoS biology *6*, 2853-2868.

d'Adda di Fagagna, F. (2008). Living on a break: cellular senescence as a DNA-damage response. Nat Rev Cancer *8*, 512-522.

d'Adda di Fagagna, F., Reaper, P.M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., Saretzki, G., Carter, N.P., and Jackson, S.P. (2003). A DNA damage checkpoint response in telomere-initiated senescence. Nature *426*, 194-198.

de Lange, T. (2005). Shelterin: the protein complex that shapes and safeguards human telomeres. Genes & development *19*, 2100-2110.

de Lange, T. (2010). How shelterin solves the telomere end-protection problem. Cold Spring Harb Symp Quant Biol *75*, 167-177.

Degerman, S., Landfors, M., Siwicki, J.K., Revie, J., Borssen, M., Evelonn, E., Forestier, E., Chrzanowska, K.H., Ryden, P., Keith, W.N.*, et al.* (2014). Immortalization of T-cells is accompanied by gradual changes in CpG methylation resulting in a profile resembling a subset of T-cell leukemias. Neoplasia *16*, 606-615.

Degerman, S., Siwicki, J.K., Osterman, P., Lafferty-Whyte, K., Keith, W.N., and Roos, G. (2010). Telomerase upregulation is a postcrisis event during senescence bypass and immortalization of two Nijmegen breakage syndrome T cell cultures. Aging Cell *9*, 220-235.

DePinho, R.A. (2000). The age of cancer. Nature *408*, 248-254.

Di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., Schurra, C., Garre, M., Nuciforo, P.G., Bensimon, A.*, et al.* (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature *444*, 638-642.

Di Micco, R., Sulli, G., Dobreva, M., Liontos, M., Botrugno, O.A., Gargiulo, G., dal Zuffo, R., Matti, V., d'Ario, G., Montani, E.*, et al.* (2011). Interplay between oncogene-induced DNA damage response and heterochromatin in senescence and cancer. Nat Cell Biol *13*, 292-302.

Dimauro, T., and David, G. (2009). Chromatin modifications: the driving force of senescence and aging? Aging (Albany NY) *1*, 182-190.

Dimri, G.P., Lee, X., Basile, G., Acosta, M., Scott, G., Roskelley, C., Medrano, E.E., Linskens, M., Rubelj, I., Pereira-Smith, O.*, et al.* (1995). A biomarker that identifies senescent human cells in culture and in aging skin in vivo. Proceedings of the National Academy of Sciences of the United States of America *92*, 9363-9367.

Durant, S.T. (2012). Telomerase-independent paths to immortality in predictable cancer subtypes. J Cancer *3*, 67-82.

Egan, E.D., and Collins, K. (2012). Biogenesis of telomerase ribonucleoproteins. RNA *18*, 1747-1759.

Ekins, S., Bugrim, A., Brovold, L., Kirillov, E., Nikolsky, Y., Rakhmatulin, E., Sorokina, S., Ryabov, A., Serebryiskaya, T., Melnikov, A.*, et al.* (2006). Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. Xenobiotica *36*, 877-901.

Feng, J., Funk, W.D., Wang, S.S., Weinrich, S.L., Avilion, A.A., Chiu, C.P., Adams, R.R., Chang, E., Allsopp, R.C., Yu, J.*, et al.* (1995). The RNA component of human telomerase. Science *269*, 1236-1241.

Flores-Rozas, H., Clark, D., and Kolodner, R.D. (2000). Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. Nature genetics *26*, 375-378.

Freund, A., Patil, C.K., and Campisi, J. (2011). p38MAPK is a novel DNA damage response-independent regulator of the senescence-associated secretory phenotype. The EMBO journal *30*, 1536-1548.

Galati, A., Magdinier, F., Colasanti, V., Bauwens, S., Pinte, S., Ricordy, R., Giraud-Panis, M.J., Pusch, M.C., Savino, M., Cacchione, S.*, et al.* (2012). TRF2 controls telomeric nucleosome organization in a cell cycle phase-dependent manner. PLoS One *7*, e34386.

Garcia-Cao, M., O'Sullivan, R., Peters, A.H., Jenuwein, T., and Blasco, M.A. (2004). Epigenetic regulation of telomere length in mammalian cells by the Suv39h1 and Suv39h2 histone methyltransferases. Nature genetics *36*, 94-99.

Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., Goesmann, A., Hibbs, M.A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D.*, et al.* (2010). Visualization of omics data for systems biology. Nat Methods *7*, S56-68.

Gil, J., and Peters, G. (2006). Regulation of the INK4b-ARF-INK4a tumour suppressor locus: all for one or one for all. Nature reviews Molecular cell biology *7*, 667-677.

Goldstein, S., Moerman, E.J., Fujii, S., and Sobel, B.E. (1994). Overexpression of plasminogen activator inhibitor type-1 in senescent fibroblasts from normal subjects and those with Werner syndrome. J Cell Physiol *161*, 571-579.

Greider, C.W. (1991). Telomerase is processive. Mol Cell Biol *11*, 4572-4580.

Hampel, S.M., Pepe, A., Greulich-Bode, K.M., Malhotra, S.V., Reszka, A.P., Veith, S., Boukamp, P., and Neidle, S. (2013). Mechanism of the antiproliferative activity of some naphthalene diimide G-quadruplex ligands. Mol Pharmacol *83*, 470-480.

Harfe, B.D., and Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. Annual review of genetics *34*, 359-399.

Harley, C.B., Futcher, A.B., and Greider, C.W. (1990). Telomeres shorten during ageing of human fibroblasts. Nature *345*, 458-460.

Harper, J.W., Adami, G.R., Wei, N., Keyomarsi, K., and Elledge, S.J. (1993). The p21 Cdk-interacting protein Cip1 is a potent inhibitor of G1 cyclin-dependent kinases. Cell *75*, 805-816.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Exp Cell Res *25*, 585-621.

He, J., Kallin, E.M., Tsukada, Y., and Zhang, Y. (2008). The H3K36 demethylase Jhdm1b/Kdm2b regulates cell proliferation and senescence through p15(Ink4b). Nature structural & molecular biology *15*, 1169-1175.

He, T.C., Zhou, S., da Costa, L.T., Yu, J., Kinzler, K.W., and Vogelstein, B. (1998). A simplified system for generating recombinant adenoviruses. Proceedings of the National Academy of Sciences of the United States of America *95*, 2509-2514.

Heidenreich, B., Rachakonda, P.S., Hemminki, K., and Kumar, R. (2014). TERT promoter mutations in cancer development. Current opinion in genetics & development *24*, 30-37.

Heiss, N.S., Knight, S.W., Vulliamy, T.J., Klauck, S.M., Wiemann, S., Mason, P.J., Poustka, A., and Dokal, I. (1998). X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nature genetics *19*, 32-38.

Herbig, U., Jobling, W.A., Chen, B.P., Chen, D.J., and Sedivy, J.M. (2004). Telomere shortening triggers senescence of human cells through a pathway involving ATM, p53, and p21(CIP1), but not p16(INK4a). Mol Cell *14*, 501-513.

Hermjakob, H., Montecchi-Palazzi, L., Lewington, C., Mudali, S., Kerrien, S., Orchard, S., Vingron, M., Roechert, B., Roepstorff, P., Valencia, A.*, et al.* (2004). IntAct: an open source molecular interaction database. Nucleic acids research *32*, D452-455.

Hernandez-Pigeon, H., Laurent, G., Humbert, O., Salles, B., and Lautier, D. (2004). Degadration of mismatch repair hMutSalpha heterodimer by the ubiquitin-proteasome pathway. FEBS Lett *562*, 40-44.

Himi, T., Yoshioka, I., and Kataura, A. (1997). Influence of age on the production of interleukin-8-like chemokine (GRO/CINC-1) in rat nasal mucosa. Eur Arch Otorhinolaryngol *254*, 101-104.

Hoenicke, L., and Zender, L. (2012). Immune surveillance of senescent cells--biological significance in cancer- and non-cancer pathologies. Carcinogenesis *33*, 1123-1126.

Holohan, B., Wright, W.E., and Shay, J.W. (2014). Cell biology of disease: Telomeropathies: an emerging spectrum disorder. The Journal of cell biology *205*, 289-299.

Hong, D.S., Angelo, L.S., and Kurzrock, R. (2007). Interleukin-6 and its receptor in cancer: implications for translational therapeutics. Cancer *110*, 1911-1928.

Horn, S., Figl, A., Rachakonda, P.S., Fischer, C., Sucker, A., Gast, A., Kadel, S., Moll, I., Nagore, E., Hemminki, K.*, et al.* (2013). TERT promoter mutations in familial and sporadic melanoma. Science *339*, 959-961.

Hsieh, P. (2001). Molecular mechanisms of DNA mismatch repair. Mutation research *486*, 71-87.

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*, 1-13.

Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L., and Garraway, L.A. (2013). Highly recurrent TERT promoter mutations in human melanoma. Science *339*, 957-959.

Isoda, T., Mitsuiki, N., Ohkawa, T., Kaneko, S., Endo, A., Ono, T., Aoki, Y., Tomizawa, D., Kajiwara, M., Araki, S.*, et al.* (2013). Irreversible leukoencephalopathy after reduced-intensity stem cell transplantation in a dyskeratosis congenita patient with TINF2 mutation. Journal of pediatric hematology/oncology *35*, e178-182.

Jackson, S.R., Zhu, C.H., Paulson, V., Watkins, L., Dikmen, Z.G., Gryaznov, S.M., Wright, W.E., and Shay, J.W. (2007). Antiadhesive effects of GRN163L--an oligonucleotide N3'->P5' thio-phosphoramidate targeting telomerase. Cancer research *67*, 1121-1129.

Jacobs, J.J., and de Lange, T. (2004). Significant role for p16INK4a in p53-independent telomere-directed senescence. Current biology : CB *14*, 2302-2308.

Jascur, T., and Boland, C.R. (2006). Structure and function of the components of the human DNA mismatch repair system. International journal of cancer Journal international du cancer *119*, 2030-2035.

Jennings, B.J., Ozanne, S.E., and Hales, C.N. (2000). Nutrition, oxidative damage, telomere shortening, and cellular senescence: individual or connected agents of aging? Mol Genet Metab *71*, 32-42.

Jiricny, J. (2006). The multifaceted mismatch-repair system. Nature reviews Molecular cell biology *7*, 335-346.

Jun, S.H., Kim, T.G., and Ban, C. (2006). DNA mismatch repair system. Classical and fresh roles. FEBS J *273*, 1609-1619.

Jurczyluk, J., Nouwens, A.S., Holien, J.K., Adams, T.E., Lovrecz, G.O., Parker, M.W., Cohen, S.B., and Bryan, T.M. (2011). Direct involvement of the TEN domain at the active site of human telomerase. Nucleic acids research *39*, 1774-1788.

Kang, T.W., Yevsa, T., Woller, N., Hoenicke, L., Wuestefeld, T., Dauch, D., Hohmeyer, A., Gereke, M., Rudalska, R., Potapova, A.*, et al.* (2011). Senescence surveillance of pre-malignant hepatocytes limits liver cancer development. Nature.

Karlseder, J., Smogorzewska, A., and de Lange, T. (2002). Senescence induced by altered telomere state, not telomere loss. Science *295*, 2446-2449.

Kennedy, A.L., McBryan, T., Enders, G.H., Johnson, F.B., Zhang, R., and Adams, P.D. (2010). Senescent mouse cells fail to overtly regulate the HIRA histone chaperone and do not form robust Senescence Associated Heterochromatin Foci. Cell Div *5*, 16.

Kida, I., Aoki, M., Ogihara, T., and Rakugi, H. (2013). Anti-apoptotic effect of human telomerase reverse transcriptase on endothelial cells under oxidative stress, independent of telomere elongation and telomerase activity. Immunology, Endocrine and Metabolic Agents in Medicinal Chemistry *13*, 112-121.

Kim, M.M., Rivera, M.A., Botchkina, I.L., Shalaby, R., Thor, A.D., and Blackburn, E.H. (2001). A low threshold level of expression of mutant-template telomerase RNA inhibits human tumor cell proliferation. Proceedings of the National Academy of Sciences of the United States of America *98*, 7982-7987.

Kim, W.Y., and Sharpless, N.E. (2006). The regulation of INK4/ARF in cancer and aging. Cell *127*, 265-275.

Kiss, T., Fayet-Lebaron, E., and Jady, B.E. (2010). Box H/ACA small ribonucleoproteins. Molecular cell *37*, 597-606.

Kitaura, H., Shinshi, M., Uchikoshi, Y., Ono, T., Iguchi-Ariga, S.M., and Ariga, H. (2000). Reciprocal regulation via protein-protein interaction between c-Myc and p21(cip1/waf1/sdi1) in DNA replication and transcription. J Biol Chem *275*, 10477-10483.

Kolodner, R.D., and Marsischky, G.T. (1999). Eukaryotic DNA mismatch repair. Current opinion in genetics & development *9*, 89-96.

Kondoh, H., Lleonart, M.E., Gil, J., Wang, J., Degan, P., Peters, G., Martinez, D., Carnero, A., and Beach, D. (2005). Glycolytic enzymes can modulate cellular life span. Cancer Res *65*, 177-185.

Kosar, M., Bartkova, J., Hubackova, S., Hodny, Z., Lukas, J., and Bartek, J. (2011). Senescence-associated heterochromatin foci are dispensable for cellular senescence, occur in a cell type- and insult-dependent manner and follow expression of p16(ink4a). Cell Cycle *10*, 457-468.

Kuilman, T., Michaloglou, C., Vredeveld, L.C., Douma, S., van Doorn, R., Desmet, C.J., Aarden, L.A., Mooi, W.J., and Peeper, D.S. (2008). Oncogene-induced senescence relayed by an interleukin-dependent inflammatory network. Cell *133*, 1019-1031.

Kuilman, T., and Peeper, D.S. (2009). Senescence-messaging secretome: SMS-ing cellular stress. Nat Rev Cancer *9*, 81-94.

Kunkel, T.A., and Erie, D.A. (2005). DNA mismatch repair. Annual review of biochemistry *74*, 681-710.

Lafferty-Whyte, K., Bilsland, A., Cairney, C.J., Hanley, L., Jamieson, N.B., Zaffaroni, N., Oien, K.A., Burns, S., Roffey, J., Boyd, S.M*., et al.* (2010a). Scoring of senescence signalling in multiple human tumour gene expression datasets, identification of a correlation between senescence score and drug toxicity in the NCI60 panel and a pro-inflammatory signature correlating with survival advantage in peritoneal mesothelioma. BMC Genomics *11*, 532.

Lafferty-Whyte, K., Bilsland, A., Hoare, S.F., Burns, S., Zaffaroni, N., Cairney, C.J., and Keith, W.N. (2010b). TCEAL7 inhibition of c-Myc activity in alternative lengthening of telomeres regulates hTERT expression. Neoplasia *12*, 405-414.

Lafferty-Whyte, K., Cairney, C.J., Will, M.B., Serakinci, N., Daidone, M.G., Zaffaroni, N., Bilsland, A., and Keith, W.N. (2009). A gene expression signature classifying telomerase and ALT immortalization reveals an hTERT regulatory network and suggests a mesenchymal stem cell origin for ALT. Oncogene *28*, 3765-3774.

Lai, C.K., Mitchell, J.R., and Collins, K. (2001). RNA binding domain of telomerase reverse transcriptase. Mol Cell Biol *21*, 990-1000.

Lau, P.J., and Kolodner, R.D. (2003). Transfer of the MSH2.MSH6 complex from proliferating cell nuclear antigen to mispaired bases in DNA. The Journal of biological chemistry *278*, 14-17.

Lawless, C., Jurk, D., Gillespie, C.S., Shanley, D., Saretzki, G., von Zglinicki, T., and Passos, J.F. (2012). A stochastic step model of replicative senescence explains ROS production rate in ageing cell populations. PLoS One *7*, e32117.

Lee, O.H., Kim, H., He, Q., Baek, H.J., Yang, D., Chen, L.Y., Liang, J., Chae, H.K., Safari, A., Liu, D*., et al.* (2011). Genome-wide YFP fluorescence complementation screen identifies new regulators for telomere signaling in human cells. Mol Cell Proteomics *10*, M110 001628.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.*, et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301-313.

Lejnine, S., Makarov, V.L., and Langmore, J.P. (1995). Conserved nucleoprotein structure at the ends of vertebrate and invertebrate chromosomes. Proc Natl Acad Sci U S A *92*, 2393-2397.

Liao, J.J. (2007). Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. Journal of medicinal chemistry *50*, 409-424.

Lin, H.K., Chen, Z., Wang, G., Nardella, C., Lee, S.W., Chan, C.H., Yang, W.L., Wang, J., Egia, A., Nakayama, K.I.*, et al.* (2010). Skp2 targeting suppresses tumorigenesis by Arf-p53-independent cellular senescence. Nature *464*, 374-379.

Lingner, J., Hughes, T.R., Shevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. Science *276*, 561-567.

Lipps, H.J., and Rhodes, D. (2009). G-quadruplex structures: in vivo evidence and function. Trends in cell biology *19*, 414-422.

Liu, D., and Hornsby, P.J. (2007). Senescent human fibroblasts increase the early growth of xenograft tumors via matrix metalloproteinase secretion. Cancer research *67*, 3117-3126.

Liu, D., Safari, A., O'Connor, M.S., Chan, D.W., Laegeler, A., Qin, J., and Songyang, Z. (2004). PTOP interacts with POT1 and regulates its localization to telomeres. Nature cell biology *6*, 673-680.

Liu, J.Q., Chen, C.Y., Xue, Y., Hao, Y.H., and Tan, Z. (2010). G-quadruplex hinders translocation of BLM helicase on DNA: a real-time fluorescence spectroscopic unwinding study and comparison with duplex substrates. Journal of the American Chemical Society *132*, 10521-10527.

Mac Partlin, M., Homer, E., Robinson, H., McCormick, C.J., Crouch, D.H., Durant, S.T., Matheson, E.C., Hall, A.G., Gillespie, D.A., and Brown, R. (2003). Interactions of the DNA mismatch repair proteins MLH1 and MSH2 with c-MYC and MAX. Oncogene *22*, 819-825.

Mahalingam, D., Tay, L.L., Tan, W.H., Chai, J.H., and Wang, X. (2011). Mutant telomerase RNAs induce DNA damage and apoptosis via the TRF2-ATM pathway in telomerase-overexpressing primary fibroblasts. FEBS J *278*, 3724-3738.

Makarov, V.L., Hirose, Y., and Langmore, J.P. (1997). Long G tails at both ends of human chromosomes suggest a C strand degradation mechanism for telomere shortening. Cell *88*, 657-666.

Mao, Z., Seluanov, A., Jiang, Y., and Gorbunova, V. (2007). TRF2 is required for repair of nontelomeric DNA double-strand breaks by homologous recombination. Proceedings of the National Academy of Sciences of the United States of America *104*, 13068-13073.

Martin-Ruiz, C., Saretzki, G., Petrie, J., Ladhoff, J., Jeyapalan, J., Wei, W., Sedivy, J., and von Zglinicki, T. (2004). Stochastic variation in telomere shortening rate causes heterogeneity of human fibroblast replicative life span. J Biol Chem *279*, 17826-17833.

Martinez, P., Siegl-Cachedenier, I., Flores, J.M., and Blasco, M.A. (2009). MSH2 deficiency abolishes the anticancer and pro-aging activity of short telomeres. Aging Cell *8*, 2-17.

Martinez, P., Thanasoula, M., Carlos, A.R., Gomez-Lopez, G., Tejera, A.M., Schoeftner, S., Dominguez, O., Pisano, D.G., Tarsounas, M., and Blasco, M.A. (2010). Mammalian Rap1 controls telomere function and gene expression through binding to telomeric and extratelomeric sites. Nature cell biology *12*, 768-780.

McCulloch, S.D., Gu, L., and Li, G.M. (2003). Bi-directional processing of DNA loops by mismatch repair-dependent and -independent pathways in human cells. The Journal of biological chemistry *278*, 3891-3896.

McGowan, J.E., Chen, L., Gao, D., Trush, M., and Wei, C. (2006). Increased mitochondrial reactive oxygen species production in newborn brain during hypoglycemia. Neurosci Lett *399*, 111-114.

Mendez-Bermudez, A., and Royle, N.J. (2011). Deficiency in DNA mismatch repair increases the rate of telomere shortening in normal human cells. Human mutation *32*, 939-946.

Menssen, A., and Hermeking, H. (2002). Characterization of the c-MYC-regulated transcriptome by SAGE: identification and analysis of c-MYC target genes. Proc Natl Acad Sci U S A *99*, 6274-6279.

Michailidi, C., Papavassiliou, A.G., and Troungos, C. (2012). DNA repair mechanisms in colorectal carcinogenesis. Current molecular medicine *12*, 237-246.

Michishita, E., McCord, R.A., Berber, E., Kioi, M., Padilla-Nash, H., Damian, M., Cheung, P., Kusumoto, R., Kawahara, T.L., Barrett, J.C.*, et al.* (2008). SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. Nature *452*, 492-496.

Minamino, T., Orimo, M., Shimizu, I., Kunieda, T., Yokoyama, M., Ito, T., Nojima, A., Nabetani, A., Oike, Y., Matsubara, H.*, et al.* (2009). A crucial role for adipose tissue p53 in the regulation of insulin resistance. Nat Med *15*, 1082-1087.

Missios, P., Zhou, Y., Guachalla, L.M., von Figura, G., Wegner, A., Chakkarappan, S.R., Binz, T., Gompf, A., Hartleben, G., Burkhalter, M.D.*, et al.* (2014). Glucose substitution prolongs maintenance of energy homeostasis and lifespan of telomere dysfunctional mice. Nature communications *5*, 4924.

Mitchell, J.R., Cheng, J., and Collins, K. (1999). A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. Mol Cell Biol *19*, 567-576.

Munoz-Espin, D., and Serrano, M. (2014). Cellular senescence: from physiology to pathology. Nature reviews Molecular cell biology *15*, 482-496.

Munoz-Pinedo, C., Ruiz-Ruiz, C., Ruiz de Almodovar, C., Palacios, C., and Lopez-Rivas, A. (2003). Inhibition of glucose metabolism sensitizes tumor cells to death receptor-triggered apoptosis through enhancement of death-inducing signaling complex formation and apical procaspase-8 processing. The Journal of biological chemistry *278*, 12759-12768.

Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. (1997). Telomerase catalytic subunit homologs from fission yeast and human. Science *277*, 955-959.

Nandakumar, J., and Cech, T.R. (2013). Finding the end: recruitment of telomerase to telomeres. Nature reviews Molecular cell biology *14*, 69-82.

Nardella, C., Clohessy, J.G., Alimonti, A., and Pandolfi, P.P. (2011). Pro-senescence therapy for cancer treatment. Nat Rev Cancer *11*, 503-511.

Narita, M., Nunez, S., Heard, E., Lin, A.W., Hearn, S.A., Spector, D.L., Hannon, G.J., and Lowe, S.W. (2003). Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence. Cell *113*, 703-716.

Nielsen, C.B., Cantor, M., Dubchak, I., Gordon, D., and Wang, T. (2010). Visualizing genomes: techniques and challenges. Nat Methods *7*, S5-S15.

Nijwening, J.H., Geutjes, E.J., Bernards, R., and Beijersbergen, R.L. (2011). The histone demethylase Jarid1b (Kdm5b) is a novel component of the Rb pathway and associates with E2f-target genes in MEFs during senescence. PLoS One *6*, e25235.

O'Connor, M.S., Safari, A., Liu, D., Qin, J., and Songyang, Z. (2004). The human Rap1 protein complex and modulation of telomere length. The Journal of biological chemistry *279*, 28585-28591.

O'Donoghue, S.I., Gavin, A.C., Gehlenborg, N., Goodsell, D.S., Heriche, J.K., Nielsen, C.B., North, C., Olson, A.J., Procter, J.B., Shattuck, D.W.*, et al.* (2010a). Visualizing biological data-now and in the future. Nat Methods *7*, S2-4.

O'Donoghue, S.I., Goodsell, D.S., Frangakis, A.S., Jossinet, F., Laskowski, R.A., Nilges, M., Saibil, H.R., Schafferhans, A., Wade, R.C., Westhof, E.*, et al.* (2010b). Visualization of macromolecular structures. Nat Methods *7*, S42-55.

O'Sullivan, R.J., Kubicek, S., Schreiber, S.L., and Karlseder, J. (2010). Reduced histone biosynthesis and chromatin changes arising from a damage signal at telomeres. Nat Struct Mol Biol *17*, 1218-1225.

Oakes, C.C., Smiraglia, D.J., Plass, C., Trasler, J.M., and Robaire, B. (2003). Aging results in hypermethylation of ribosomal DNA in sperm and liver of male rats. Proceedings of the National Academy of Sciences of the United States of America *100*, 1775-1780.

Ogryzko, V.V., Hirai, T.H., Russanova, V.R., Barbie, D.A., and Howard, B.H. (1996). Human fibroblast commitment to a senescence-like state in response to histone deacetylase inhibitors is cell cycle dependent. Mol Cell Biol *16*, 5210-5218.

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*, 763-774.

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

Panero, J., O'Callaghan, N.J., Fenech, M., and Slavutsky, I. (2015). Absolute qPCR for measuring telomere length in bone marrow samples of plasma cell disorders. Mol Biotechnol *57*, 155-159.

Parkinson, H., Kapushesky, M., Shojatalab, M., Abeygunawardena, N., Coulson, R., Farne, A., Holloway, E., Kolesnykov, N., Lilja, P., Lukk, M.*, et al.* (2007). ArrayExpress--a public database of microarray experiments and gene expression profiles. Nucleic acids research *35*, D747-750.

Passos, J.F., Nelson, G., Wang, C., Richter, T., Simillion, C., Proctor, C.J., Miwa, S., Olijslagers, S., Hallinan, J., Wipat, A.*, et al.* (2010). Feedback between p21 and reactive oxygen production is necessary for cell senescence. Mol Syst Biol *6*, 347.

Passos, J.F., Saretzki, G., Ahmed, S., Nelson, G., Richter, T., Peters, H., Wappler, I., Birket, M.J., Harold, G., Schaeuble, K.*, et al.* (2007). Mitochondrial dysfunction accounts for the stochastic heterogeneity in telomere-dependent senescence. PLoS Biol *5*, e110.

Pazolli, E., Alspach, E., Milczarek, A., Prior, J., Piwnica-Worms, D., and Stewart, S.A. (2012). Chromatin remodeling underlies the senescence-associated secretory phenotype of tumor stromal fibroblasts that supports cancer progression. Cancer Res *72*, 2251-2261.

Prieur, A., and Peeper, D.S. (2008). Cellular senescence in vivo: a barrier to tumorigenesis. Curr Opin Cell Biol *20*, 150-155.

Procter, J.B., Thompson, J., Letunic, I., Creevey, C., Jossinet, F., and Barton, G.J. (2010). Visualization of multiple alignments, phylogenies and gene family evolution. Nat Methods *7*, S16-25.

Rai, T.S., and Adams, P.D. (2012). Lessons from senescence: Chromatin maintenance in non-proliferating cells. Biochim Biophys Acta *1819*, 322-331.

Robart, A.R., and Collins, K. (2011). Human telomerase domain interactions capture DNA for TEN domain-dependent processive elongation. Molecular cell *42*, 308-318.

Rodier, F., Coppe, J.P., Patil, C.K., Hoeijmakers, W.A., Munoz, D.P., Raza, S.R., Freund, A., Campeau, E., Davalos, A.R., and Campisi, J. (2009). Persistent DNA damage signalling triggers senescence-associated inflammatory cytokine secretion. Nat Cell Biol *11*, 973-979.

Rodier, F., Munoz, D.P., Teachenor, R., Chu, V., Le, O., Bhaumik, D., Coppe, J.P., Campeau, E., Beausejour, C.M., Kim, S.H.*, et al.* (2011). DNA-SCARS: distinct nuclear structures that sustain damage-induced senescence growth arrest and inflammatory cytokine secretion. J Cell Sci *124*, 68-81.

Rudolph, K.L., Millard, M., Bosenberg, M.W., and DePinho, R.A. (2001). Telomere dysfunction and evolution of intestinal carcinoma in mice and humans. Nature genetics *28*, 155-159.

Sakuraba, H., Mizukami, H., Yagihashi, N., Wada, R., Hanyu, C., and Yagihashi, S. (2002). Reduced beta-cell mass and expression of oxidative stress-related DNA damage in the islet of Japanese Type II diabetic patients. Diabetologia *45*, 85-96.

Salama, R., Sadaie, M., Hoare, M., and Narita, M. (2014). Cellular senescence and its effector programs. Genes & development *28*, 99-114.

Sasa, G.S., Ribes-Zamora, A., Nelson, N.D., and Bertuch, A.A. (2012). Three novel truncating TINF2 mutations causing severe dyskeratosis congenita in early childhood. Clin Genet *81*, 470-478.

Savage, S.A., Giri, N., Baerlocher, G.M., Orr, N., Lansdorp, P.M., and Alter, B.P. (2008). TINF2, a component of the shelterin telomere protection complex, is mutated in dyskeratosis congenita. Am J Hum Genet *82*, 501-509.

Schmidt, J.C., Dalby, A.B., and Cech, T.R. (2014). Identification of human TERT elements necessary for telomerase recruitment to telomeres. eLife *3*.

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*, 131-138.

Shaheen, F., Grammatopoulos, D.K., Muller, J., Zammit, V.A., and Lehnert, H. (2014). Extra-nuclear telomerase reverse transcriptase (TERT) regulates glucose transport in skeletal muscle cells. Biochim Biophys Acta *1842*, 1762-1769.

Shay, J.W., and Wright, W.E. (1996). Telomerase activity in human cancer. Curr Opin Oncol *8*, 66-71.

Shay, J.W., and Wright, W.E. (2011). Role of telomeres and telomerase in cancer. Seminars in cancer biology *21*, 349-353.

Siegl-Cachedenier, I., Munoz, P., Flores, J.M., Klatt, P., and Blasco, M.A. (2007). Deficient mismatch repair improves organismal fitness and survival of mice with dysfunctional telomeres. Genes & development *21*, 2234-2247.

Simonet, T., Zaragosi, L.E., Philippe, C., Lebrigand, K., Schouteden, C., Augereau, A., Bauwens, S., Ye, J., Santagostino, M., Giulotto, E.*, et al.* (2011). The human TTAGGG repeat factors 1 and 2 bind to a subset of interstitial telomeric sequences and satellite repeats. Cell Res *21*, 1028-1038.

Singh, K., Matsuyama, S., Drazba, J.A., and Almasan, A. (2012). Autophagy-dependent senescence in response to DNA damage and chronic apoptotic stress. Autophagy *8*, 236-251.

Smogorzewska, A., and de Lange, T. (2004). Regulation of telomerase by telomeric proteins. Annual review of biochemistry *73*, 177-208.

Sone, H., and Kagawa, Y. (2005). Pancreatic beta cell senescence contributes to the pathogenesis of type 2 diabetes in high-fat diet-induced diabetic mice. Diabetologia *48*, 58-67.

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

Tamura, Y., Izumiyama-Shimomura, N., Kimbara, Y., Nakamura, K., Ishikawa, N., Aida, J., Chiba, Y., Mori, S., Arai, T., Aizawa, T.*, et al.* (2014). beta-cell telomere attrition in diabetes: inverse correlation between HbA1c and telomere length. J Clin Endocrinol Metab *99*, 2771-2777.

Tang, J., Kan, Z.Y., Yao, Y., Wang, Q., Hao, Y.H., and Tan, Z. (2008). G-quadruplex preferentially forms at the very 3' end of vertebrate telomeric DNA. Nucleic acids research *36*, 1200-1208.

Tejera, A.M., Stagno d'Alcontres, M., Thanasoula, M., Marion, R.M., Martinez, P., Liao, C., Flores, J.M., Tarsounas, M., and Blasco, M.A. (2010). TPP1 is required for TERT recruitment, telomere elongation during nuclear reprogramming, and normal skin development in mice. Dev Cell *18*, 775-789.

Teo, H., Ghosh, S., Luesch, H., Ghosh, A., Wong, E.T., Malik, N., Orth, A., de Jesus, P., Perry, A.S., Oliver, J.D.*, et al.* (2010). Telomere-independent Rap1 is an IKK adaptor and regulates NF-kappaB-dependent gene expression. Nature cell biology *12*, 758-767.

Thanasoula, M., Escandell, J.M., Martinez, P., Badie, S., Munoz, P., Blasco, M.A., and Tarsounas, M. (2010). p53 prevents entry into mitosis with uncapped telomeres. Current biology : CB *20*, 521-526.

Thanasoula, M., Escandell, J.M., Suwaki, N., and Tarsounas, M. (2012a). ATM/ATR checkpoint activation downregulates CDC25C to prevent mitotic entry with uncapped telomeres. The EMBO journal.

Thanasoula, M., Escandell, J.M., Suwaki, N., and Tarsounas, M. (2012b). ATM/ATR checkpoint activation downregulates CDC25C to prevent mitotic entry with uncapped telomeres. The EMBO journal *31*, 3398-3410.

Theimer, C.A., and Feigon, J. (2006). Structure and function of telomerase RNA. Current opinion in structural biology *16*, 307-318.

Tong, X.J., Li, Q.J., Duan, Y.M., Liu, N.N., Zhang, M.L., and Zhou, J.Q. (2011). Est1 protects telomeres and inhibits subtelomeric y'-element recombination. Mol Cell Biol *31*, 1263-1274.

Touzot, F., Callebaut, I., Soulier, J., Gaillard, L., Azerrad, C., Durandy, A., Fischer, A., de Villartay, J.P., and Revy, P. (2010). Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. Proceedings of the National Academy of Sciences of the United States of America *107*, 10097-10102.

Tran, P.T., Simon, J.A., and Liskay, R.M. (2001). Interactions of Exo1p with components of MutLalpha in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America *98*, 9760-9765.

Uziel, O., Singer, J.A., Danicek, V., Sahar, G., Berkov, E., Luchansky, M., Fraser, A., Ram, R., and Lahav, M. (2007). Telomere dynamics in arteries and mononuclear cells of diabetic patients: effect of diabetes and of glycemic control. Exp Gerontol *42*, 971-978.

Vaish, M., and Mittal, B. (2002). DNA mismatch repair, microsatellite instability and cancer. Indian journal of experimental biology *40*, 989-994.

Van Den Steen, P.E., Wuyts, A., Husson, S.J., Proost, P., Van Damme, J., and Opdenakker, G. (2003). Gelatinase B/MMP-9 and neutrophil collagenase/MMP-8 process the chemokines human GCP-2/CXCL6, ENA-78/CXCL5 and mouse GCP-2/LIX and modulate their physiological activities. Eur J Biochem *270*, 3739-3749.

van Steensel, B., Smogorzewska, A., and de Lange, T. (1998). TRF2 protects human telomeres from end-to-end fusions. Cell *92*, 401-413.

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*, 795-806.

Ventura, A., Kirsch, D.G., McLaughlin, M.E., Tuveson, D.A., Grimm, J., Lintault, L., Newman, J., Reczek, E.E., Weissleder, R., and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. Nature *445*, 661-665.

Vilar, E., and Gruber, S.B. (2010). Microsatellite instability in colorectal cancer-the stable evidence. Nature reviews Clinical oncology *7*, 153-162.

von Kriegsheim, A., Baiocchi, D., Birtwistle, M., Sumpton, D., Bienvenut, W., Morrice, N., Yamada, K., Lamond, A., Kalna, G., Orton, R.*, et al.* (2009). Cell fate decisions are specified by the dynamic ERK interactome. Nat Cell Biol *11*, 1458-1464.

Vulliamy, T., Beswick, R., Kirwan, M., Marrone, A., Digweed, M., Walne, A., and Dokal, I. (2008). Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. Proceedings of the National Academy of Sciences of the United States of America *105*, 8073-8078.

Vulliamy, T., Beswick, R., Kirwan, M.J., Hossain, U., Walne, A.J., and Dokal, I. (2012). Telomere length measurement can distinguish pathogenic from non-pathogenic variants in the shelterin component, TIN2. Clin Genet *81*, 76-81.

Vulliamy, T., Marrone, A., Goldman, F., Dearlove, A., Bessler, M., Mason, P.J., and Dokal, I. (2001). The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. Nature *413*, 432-435.

Walne, A.J., Vulliamy, T., Marrone, A., Beswick, R., Kirwan, M., Masunari, Y., Al-Qurashi, F.H., Aljurf, M., and Dokal, I. (2007). Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. Human molecular genetics *16*, 1619-1629.

Wang, C., Mayer, J.A., Mazumdar, A., Fertuck, K., Kim, H., Brown, M., and Brown, P.H. (2011a). Estrogen induces c-myc gene expression via an upstream enhancer activated by the estrogen receptor and the AP-1 transcription factor. Mol Endocrinol *25*, 1527-1538.

Wang, D., and Jang, D.J. (2009). Protein kinase CK2 regulates cytoskeletal reorganization during ionizing radiation-induced senescence of human mesenchymal stem cells. Cancer Res *69*, 8200-8207.

Wang, H., Larris, B., Peiris, T.H., Zhang, L., Le Lay, J., Gao, Y., and Greenbaum, L.E. (2007). C/EBPbeta activates E2F-regulated genes in vivo via recruitment of the coactivator CREB-binding protein/P300. J Biol Chem *282*, 24679-24688.

Wang, Q., Liu, J.Q., Chen, Z., Zheng, K.W., Chen, C.Y., Hao, Y.H., and Tan, Z. (2011b). G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. Nucleic acids research *39*, 6229-6237.

Wardi, L., Alaaeddine, N., Raad, I., Sarkis, R., Serhal, R., Khalil, C., and Hilal, G. (2014). Glucose restriction decreases telomerase activity and enhances its inhibitor response on breast cancer cells: possible extra-telomerase role of BIBR 1532. Cancer cell international *14*, 60.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189-1191.

Wright, W.E., Piatyszek, M.A., Rainey, W.E., Byrd, W., and Shay, J.W. (1996). Telomerase activity in human germline and embryonic tissues and cells. Dev Genet *18*, 173-179.

Wu, P., Takai, H., and de Lange, T. (2012). Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST. Cell *150*, 39-52.

Wu, R.A., and Collins, K. (2014). Human telomerase specialization for repeat synthesis by unique handling of primer-template duplex. The EMBO journal *33*, 921-935.

Xue, W., Zender, L., Miething, C., Dickins, R.A., Hernando, E., Krizhanovsky, V., Cordon-Cardo, C., and Lowe, S.W. (2007). Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. Nature *445*, 656-660.

Yamaguchi, H., Calado, R.T., Ly, H., Kajigaya, S., Baerlocher, G.M., Chanock, S.J., Lansdorp, P.M., and Young, N.S. (2005). Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia. The New England journal of medicine *352*, 1413-1424.

Yang, D., Xiong, Y., Kim, H., He, Q., Li, Y., Chen, R., and Songyang, Z. (2011). Human telomeric proteins occupy selective interstitial sites. Cell Res *21*, 1013-1027.

Yang, Q., Zheng, Y.L., and Harris, C.C. (2005). POT1 and TRF2 cooperate to maintain telomeric integrity. Mol Cell Biol *25*, 1070-1080.

Yaswen, P., MacKenzie, K.L., Keith, W.N., Hentosh, P., Rodier, F., Zhu, J., Firestone, G.L., Matheu, A., Carnero, A., Bilsland, A.*, et al.* (2015). Therapeutic targeting of replicative immortality. Seminars in cancer biology.

Ye, J.Z., Donigian, J.R., van Overbeek, M., Loayza, D., Luo, Y., Krutchinsky, A.N., Chait, B.T., and de Lange, T. (2004). TIN2 binds TRF1 and TRF2 simultaneously and stabilizes the TRF2 complex on telomeres. The Journal of biological chemistry *279*, 47264-47271.

Young, A.R., and Narita, M. (2009). SASP reflects senescence. EMBO Rep *10*, 228-230.

Young, N.S. (2012). Bone marrow failure and the new telomere diseases: practice and research. Hematology *17 Suppl 1*, S18-21.

Zahler, A.M., Williamson, J.R., Cech, T.R., and Prescott, D.M. (1991). Inhibition of telomerase by G-quartet DNA structures. Nature *350*, 718-720.

Zhang, M.L., Tong, X.J., Fu, X.H., Zhou, B.O., Wang, J., Liao, X.H., Li, Q.J., Shen, N., Ding, J., and Zhou, J.Q. (2010). Yeast telomerase subunit Est1p has guanine quadruplex-promoting activity that is required for telomere elongation. Nature structural & molecular biology *17*, 202-209.

Zhang, P., Pazin, M.J., Schwartz, C.M., Becker, K.G., Wersto, R.P., Dilley, C.M., and Mattson, M.P. (2008). Nontelomeric TRF2-REST interaction modulates neuronal gene silencing and fate of tumor and stem cells. Current biology : CB *18*, 1489-1494.

Zhang, Q., Kim, N.K., and Feigon, J. (2011). Architecture of human telomerase RNA. Proceedings of the National Academy of Sciences of the United States of America *108*, 20325-20332.

Zhang, R., Chen, W., and Adams, P.D. (2007). Molecular dissection of formation of senescence-associated heterochromatin foci. Mol Cell Biol *27*, 2343-2358.

Zhang, R., Poustovoitov, M.V., Ye, X., Santos, H.A., Chen, W., Daganzo, S.M., Erzberger, J.P., Serebriiskii, I.G., Canutescu, A.A., Dunbrack, R.L.*, et al.* (2005). Formation of MacroH2A-containing senescence-associated heterochromatin foci and senescence driven by ASF1a and HIRA. Dev Cell *8*, 19-30.

Zhong, F., Savage, S.A., Shkreli, M., Giri, N., Jessop, L., Myers, T., Chen, R., Alter, B.P., and Artandi, S.E. (2011). Disruption of telomerase trafficking by TCAB1 mutation causes dyskeratosis congenita. Genes & development *25*, 11-16.

Publications

Bilsland, A.E., Pugliese, A., Liu, Y., **Revie, J.**, Burns, S., McCormick, C., Cairney, C.J., Bower, J., Drysdale, M., Narita, M., et al. (2015). Identification of a Selective G1-Phase Benzimidazolone Inhibitor by a Senescence-Targeted Virtual Screen Using Artificial Neural Networks. *Neoplasia* 17, 704-715.

Degerman, S., Landfors, M., Siwicki, J.K., **Revie, J.**, Borssen, M., Evelonn, E., Forestier, E., Chrzanowska, K.H., Ryden, P., Keith, W.N., et al. (2014). Immortalization of T-cells is accompanied by gradual changes in CpG methylation resulting in a profile resembling a subset of T-cell leukemias. *Neoplasia* 16, 606-615.

NEOPLASIA
www.neoplasia.com

Volume 17 Number 9 September 2015 pp. 704–715 704

Identification of a Selective G1-Phase Benzimidazolone Inhibitor by a Senescence-Targeted Virtual Screen Using Artificial Neural Networks[1,2]

Alan E. Bilsland*, Angelo Pugliese†, Yu Liu*, John Revie*, Sharon Burns*, Carol McCormick*, Claire J. Cairney*, Justin Bower†, Martin Drysdale†, Masashi Narita‡, Mahito Sadaie§ and W. Nicol Keith*

*Institute of Cancer Sciences, University of Glasgow, Wolfson Wohl Cancer Research Centre, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1QH, UK; †Cancer Research UK Beatson Institute, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK; ‡University of Cambridge, Cancer Research UK Cambridge Institute, Li Ka Shing Centre Robinson Way, Cambridge, CB2 0RE, UK; §Kyoto University, Graduate School of Biostudies, Yoshidakonoe-cho, Sakyo-ku Kyoto 606-8501 Japan

Abstract

Cellular senescence is a barrier to tumorigenesis in normal cells, and tumor cells undergo senescence responses to genotoxic stimuli, which is a potential target phenotype for cancer therapy. However, in this setting, mixed-mode responses are common with apoptosis the dominant effect. Hence, more selective senescence inducers are required. Here we report a machine learning–based *in silico* screen to identify potential senescence agonists. We built profiles of differentially affected biological process networks from expression data obtained under induced telomere dysfunction conditions in colorectal cancer cells and matched these to a panel of 17 protein targets with confirmatory screening data in PubChem. We trained a neural network using 3517 compounds identified as active or inactive against these targets. The resulting classification model was used to screen a virtual library of ~2M lead-like compounds. One hundred and forty-seven virtual hits were acquired for validation in growth inhibition and senescence-associated β-galactosidase assays. Among the found hits, a benzimidazolone compound, CB-20903630, had low micromolar IC50 for growth inhibition of HCT116 cells and selectively induced senescence-associated β-galactosidase activity in the entire treated cell population without cytotoxicity or apoptosis induction. Growth suppression was mediated by G1 blockade involving increased p21 expression and suppressed cyclin B1, CDK1, and CDC25C. In addition, the compound inhibited growth of multicellular spheroids and caused severe retardation of population kinetics in long-term treatments. Preliminary structure-activity and structure clustering analyses are reported, and expression analysis of CB-20903630 against other cell cycle suppressor compounds suggested a PI3K/AKT-inhibitor–like profile in normal cells, with different pathways affected in cancer cells.

Neoplasia (2015) 17, 704–715

Address all correspondence to: W. Nicol Keith, Institute of Cancer Sciences, University of Glasgow, Wolfson Wohl Cancer Research Centre, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1QH, UK.
E-mail: nicol.keith@glasgow.ac.uk

[1] This work was supported by a Development Fund award from the Cancer Research UK Glasgow Centre, the Glasgow Experimental Cancer Medicine Centre (funded by Cancer Research UK and the Chief Scientist Office, Scotland), Cancer Research UK grants C2193/A15584 and C301/A12962, University of Cambridge, Cancer Research UK Cambridge Institute Core Grant, and Hutchison Whampoa (M.N.).
[2] Conflict of Interest Statement: The authors have nothing to disclose.
Received 2 April 2015; Revised 28 August 2015; Accepted 31 August 2015

1476-5586
http://dx.doi.org/10.1016/j.neo.2015.08.009

Introduction

Cellular senescence in normal cells is an irreversible cell cycle arrest which is involved in cellular aging and tissue maintenance, and which is induced by critically shortened telomeres at the end of replicative lifespan. Oxidative damage and oncogene activation accelerate both telomere shortening and senescence induction [1]. Therefore, senescence is considered to be a barrier to tumorigenesis which cancer cells must bypass to acquire a transformed phenotype [2,3].

Many cancer cells retain the capacity to undergo senescence-like growth arrest in response to agents including chemotherapeutics and ionizing radiation in addition to many targeted agents [4]. Hence, despite inactivation of some key pathways, many tumor cells retain the ability to exit the cell cycle under appropriate treatments. Thus, latent senescence signaling may persist in tumors [5].

There is substantial interest in senescence induction as a therapeutic outcome in cancer. However, senescence involves multiple processes including telomere homeostasis, DNA damage and inflammatory signaling, chromatin regulation, and metabolism [6,7]. Interaction of these with the diverse mutational backgrounds of cancer cells adds further complexity in attempting to define the best targets for therapeutic intervention. It seems likely that a spectrum of senescence-like responses is possible in cancer cells depending on induction agent and signaling environment [8,9].

Given limitations in current knowledge, phenotypic screening is attractive both for compound and pathway discovery focused on senescence [10–12]. Suitable phenotypic markers for assay development include p21 and p16 levels, the senescence-associated secretory phenotype, senescence-associated β-galactosidase (SA-β-gal) staining, senescence-associated heterochromatin foci, and altered morphology [1]. However, although many agents elicit senescence, responses obtained are often restricted to subsets of cells, with apoptotic cell death dominant [13].

To evaluate senescence induction as an anticancer modality will require identification of senescence agonists which are substantially more selective than currently available tools [14]. Without detailed knowledge of targets, the screening challenge is not simply identification of compounds which can cause senescence; rather, stratification of the most selective compounds among many expected partial actives is critical. Identification of enriched libraries would be beneficial before initiating a screening campaign. We reasoned that virtual screening might identify such an enriched set.

Ligand-based virtual screening is of increasing interest in the construction of activity models, ranging from well-defined target binding studies [15] to more complex scenarios such as modeling of experimental microsomal stability results [16], and a wide variety of platforms and datasets are now available [17]. Another major goal is to identify new compounds with activity against a given target based on feature recognition [18].

In either case, abstraction of chemical structure information into a set of numerical descriptors is critical. These must provide detailed representation of the chemical and property space for a given compound set [19]. An assumption is that a relation can be made between these "fingerprints" and a classifier (active/inactive) or known quantity such as IC50. Machine learning methods such as neural networks [18,20] or support vector machines [21,22] provide a powerful approach. Feature recognition rules are learned from a training set with known activity; trained models are then simulated against a new compound set of unknown activity.

Here we report a virtual screen using an artificial neural network ensemble trained by the scaled conjugate gradient descent method [23] using compounds identified from pooled PubChem screens [24,25] against a panel of senescence-related targets. Targets were selected by matching available screens to cellular "process networks profiles" obtained by functional enrichment analysis of expression data in colorectal cancer cells with induced telomere dysfunction. The trained ensemble was used to classify a library of around 2M lead-like compounds, leading to identification of a benzimidazolone compound with low micromolar IC50 which selectively induces G1 blockade and SA-β-gal without causing apoptosis. Preliminary structure/activity relationships (SARs) and clustering studies are reported.

Results

Identification of a Senescence-Associated Protein Target Panel

Ad-hTR-mut is an adenoviral vector harboring mutant telomere template sequence [2,26]. Telomerase-dependent reverse transcription in cancer cells incorporates mutant sequence in the telomeres of infected cells, causing rapid telomere damage signaling. This provides a highly selective way to induce telomere dysfunction and cellular senescence.

To identify pathways associated with telomere dysfunction and senescence, we performed expression profiling and pathway analysis [27–29] on HCT116 colorectal cancer cells infected with Ad-hTR-mut or treated in long-term culture with telomerase inhibitor GRN163L [30,31]. "Process network profiles" were generated by pathway enrichment analysis against 169 curated networks from the MetaCore database [27] covering 23 top-level processes (Supplementary Table S1). Heat maps were generated based on significance of each network to visualize overall significance of each process (Figure 1*A* and Supplementary Figure S1). The most significant enrichments of differentially expressed genes in response to telomere targeting agents were on networks involved in DNA damage, cell cycle, and protein folding.

We hypothesized that targets involved in these telomere dysfunction processes would be good candidate targets for senescence induction. We therefore sought to identify a target group with known involvement in these processes and for which confirmatory (dose-response) screening results were available within the PubChem bioassay database [24,25]. In searching available screens, we identified 17 candidate protein targets with relevance to these processes and with associated confirmatory screens (Table 1).

Enrichment analysis on this target list confirmed close involvement of the panel in the same process networks identified as significantly affected by telomere targeting (Figure 1*A* and Supplementary Figure S1). We also performed shortest-paths analysis of the target panel in MetaCore to determine functional relations between these targets (Figure 1*B*). These targets participate in a closely connected direct interactions network, indicating the close interplay between diverse processes in senescence regulation.

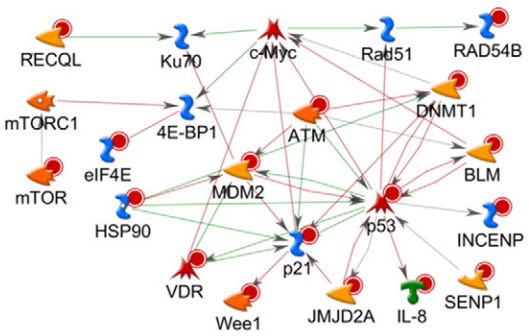
Development of a Senescence-Targeted Virtual Screen

The overall neural network optimization workflow is shown in Supplementary Figure S2. To develop the classifier, compound sets associated with each identified PubChem bioassay were merged into active and inactive pools (Table 1). As classification models can be affected by an initial unbalance in the data, we aimed at retaining similar compound numbers in each list. Similarity filters were applied to reduce the size of very large inactive lists using ChemOffice. The pooled lists were cleaned, duplicates were removed, and a molecular

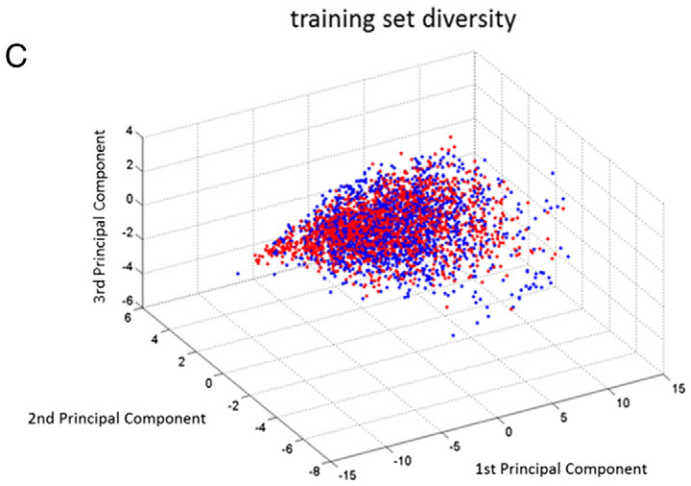
A

Process	Network	Pubchem target profile	HCT116	
			GRN163L	hTR-mut
DNA damage	BER-NER	0.03	0.04	0.08
	Core	0.00	0.27	0.76
	DSB repair	0.00	0.00	0.05
	MMR	0.00	0.07	0.00
	Checkpoint	0.00	0.00	0.00
Cell cycle	Proteolysis (cell cycle)	1.00	0.99	0.25
	Spindle microtubules	0.25	0.00	0.00
	G1-S growth factor regulation	0.09	0.16	0.01
	Mitosis	0.08	0.00	0.00
	Core	0.00	0.00	0.00
	G1-S interleukin regulation	0.00	0.10	0.16
	Meiosis	0.00	0.00	0.00
	G0-G1	0.00	0.28	0.25
	S-phase	0.00	0.00	0.00
	G1-S	0.00	0.00	0.00
	G2-M	0.00	0.00	0.00
Protein folding	Protein folding (nucleus)	0.14	0.00	0.00
	Folding in normal condition	0.04	0.00	0.51
	Response to unfolded proteins	0.01	0.01	0.65
	ER and cytoplasm	0.01	0.02	0.17

B



C



D

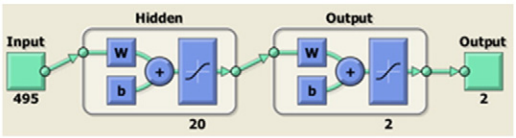


Figure 1. Development of a senescence-targeted virtual screen. (A) Expression microarray data from colorectal cancer cells were used to generate process network profiles of induced telomere dysfunction. Top scoring processes are shown. The complete profile is given in Supplementary Figure S1. Processes in red scored as significant. Numbers in each cell are the hypergeometric *P* value for each gene list against each process network. The main processes affected were matched to targets involved in those processes for which confirmatory screens were available in PubChem. (B) Interactions in the senescence-associated target panel network identified by direct-interactions network building in GeneGo. (C) Diversity of the 3517-compound training set. Principal component analysis was performed on the 495 selected chemical descriptors, and projections on the first three principal components were visualized in Matlab. Actives are shown in blue; inactives are shown in red. (D) Structure of the trained networks. A 10-network ensemble was used for the virtual screen.

weight filter was applied (150 < MW < 700). The actives and inactives were then merged to generate the training set, and duplicates arising after the merge were excluded. The more balanced training set contained 3924 compounds of which 1859 were active against the targets panel and 2065 were inactive.

For each compound, 2780 descriptors were generated. These included 729 1D/2D descriptors, 880 PubChem fingerprints and 1024 CDK extended fingerprints obtained using PaDel, and 147 pharmacophore fingerprints obtained using PowerMV [32,33]. An initial parameter scan was performed on this training set to identify the classifier performance when learning rule and neuron number were varied (Supplementary Figure S3). The best initial parameter set (20 neurons, 1 hidden layer, with learning by scaled conjugate descent) gave 73.8% classification accuracy with Cohen's $\kappa = 0.51$ [34] in 10-fold cross-validation. To improve the performance, we excluded the compounds that were most consistently misclassified under these parameters (Mean Square Error > 0.4 excluded). A total of 3536 compounds were retained, and a features selection protocol (see methods) was performed on the descriptor set for these. Nineteen compounds were not correctly recognized by the feature selection software and were excluded.

We retained 495 descriptors for the final 3517 compound set. Principal component analysis on these descriptors is shown in Figure 1*C*, illustrating good overlap between the active and inactive

Table 1. Selected PubChem Bioassay Target Panel and Associated Compounds Identified as Relevant to Telomere-Dysfunction Process Network Profiles Generated in HCT116 Cells

Target	PubChem AID	Actives	Inactives	Description
p21	N/A	29	29	In-house screening data. Luciferase assay for activation of p21 promoter activity; inactives 50%-85% similarity with actives.
p53	624305	296	405	Confirmatory luciferase assay for activation of p53-dependent synthetic promoter reporter.
WEE1	1410	39	147	Increased WEE1-luciferase fusion gene activity; inactives 65% similarity to actives.
INCENP	473665	8	0	Small series of aurora inhibitors based on modification of an existing clinical candidate.
IL8	651758	38	88	Time-resolved FRET assay (IF) for IL8 secretion from cells; inactives 65% similarity to actives.
ATM	493192	41	36	Confirmatory ELISA for phosphorylation of ATM target protein.
MTORC1	2668	49	0	Confirmatory cell-based IF assay for phospho-rpS6.
HSP90	712	91	173	Confirmatory FP assay for HSP90 binding.
DNMT1	602386	179	21	Confirmatory fluorescein-labeled DNA oligomethylation assay.
BLM	2585	83	55	Confirmatory fluorescence quench DNA unwinding assay.
MDM2	1394	41	159	Confirmatory MDM2-luc autoubiquitination assay.
RECQL1	2708	173	321	Confirmatory fluorescence quench DNA unwinding assay.
SENP1	651697	117	60	Confirmatory kinetic FRET assay for SENP protease inhibition.
VDR	602201	159	115	Confirmatory FP assay for interaction of VDR and coregulator peptide.
EIF4E	855	77	486	Confirmatory TR-FRET for association of EIF4E/EIF4G.
RAD54	651657	394	63	Confirmatory fluorescent HR assay.
JMJD2A	488840	43	0	Confirmatory dissociation enhanced lanthanide fluorescence assay.

WEE1, homologue of S.Pombe Wee1; INCENP, Inner Centromere Protein; IL8, Interleukin 8; ATM, Ataxia Telangiectasia Mutated; MTORC1, Mammalian Target of Rapamycin Complex 1; HSP90, Heat Shock Protein (90kDa); DNMT1, DNA Methyl Transferase 1; BLM, Bloom Syndrome; MDM2, Mouse Double Minute 2 homologue; RECQL1, E.Coli RecQ Like helicase 1; SENP1, Sentrin Specific Protease family member 1; VDR, Vitamin D Receptor; EIF4E, Eukaryotic Translation Initiation Factor 4E; RAD54, homologue of S.Cerevisiae Rad54; JMJD2A, Jumanji Domain containing protein 2A; FRET, Fluorescence Resonance Energy Transfer; IF, immunofluorescence; ELISA, Enzyme Linked Immuno-Sorbent Assay; rpS6, Small Ribosomal Protein 6; FP, Fluorescence Polarization; TR-FRET, Time Resolved FRET; HR, Homologous Recombination.

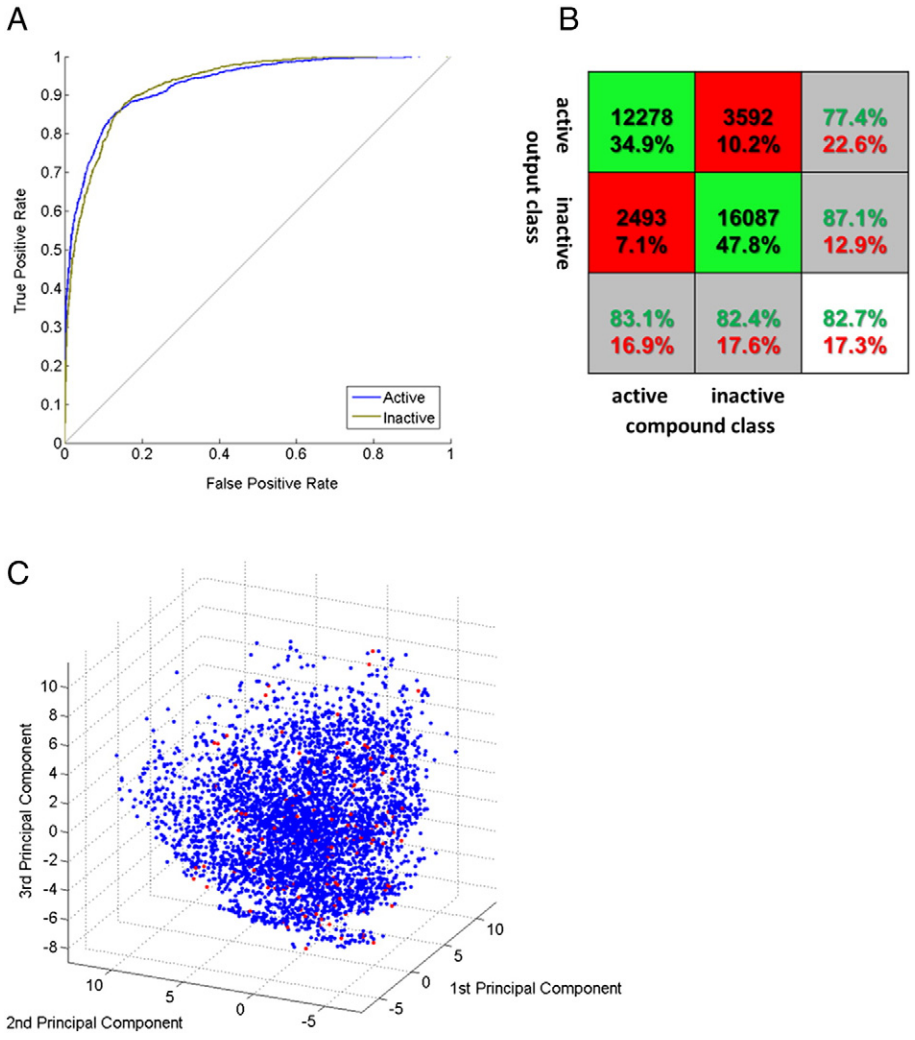


Figure 2. Performance of the network ensemble and virtual screening results. (A) Receiver operating characteristic plot of the performance of 1 of the 10 networks in the trained ensemble showing results for each output neuron. One neuron each classified active or inactive compounds. (B) Summed confusion matrix for the 10-network classifier. Numbers represent total compound number and percentage of the training set falling in each quadrant as classified across all networks. Cohen's $\kappa = 0.65$ for the ensemble. (C) Principal component analysis of filtered virtual screening hits (total set in blue) and compounds selected after clustering on 3D pharmacophores (red). Principal component analysis was performed in Matlab on 3D pharmacophores extracted using Canvas.

chemical spaces. Parameter scanning was again performed on this data with 10-fold cross-validation. Best accuracy (Supplementary Figure S4) was obtained with 2-layer networks having 20 neurons in the hidden layer with 2 output neurons (Figure 1*D*) trained by scaled conjugate gradient descent [23]. We finally selected an ensemble of 10 networks trained using the optimal parameter, compound, and descriptor sets. A representative receiver operating characteristic plot for one of these is shown in Figure 2*A*. The high area under the curve of each output neuron indicates excellent classification of both active and inactive compounds. The overall sensitivity and specificity of the entire panel were 83.1% and 82.4%, respectively, for the "hit" output neuron (Figure 2*B*). Overall accuracy was 82.7%, and Cohen's κ = 0.65 indicated very good classification performance [35].

We used the trained ensemble to screen 2,086,587 structures with weighting for specificity by application of a cutoff of 0.95 on the active output neuron and 0.05 on the inactive output neuron, resulting in 17,278 virtual hits. To prioritize these, *in silico* physicochemical and ADMET filters were applied (Supplementary Table S2). A total of 4929 compounds remained after filtering, and these were clustered on 3D pharmacophore fingerprints. A final set of 147 for cell-based screening was obtained by sampling from these clusters (Figure 2*C*; SMILES structures are given in Supplementary File 1).

Identification of a Highly Selective Benzimidazolone SA-β-gal Inducer

The 147 compounds were first tested in MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide) cell viability assays in HCT116 cells at 100 μM. Ninety-two compounds showed at least 1.5-fold growth inhibition and were taken into dose-response treatments (Supplementary Figure S4). Sixty of the 92 compounds showed confirmed growth inhibition with IC50 < 100 μM (Figure 3*A*), and the top 50 of these with >60% growth suppression at 100 μM were tested in a fluorometric SA-β-gal assay. Each compound was tested initially at a single dose determined from the MTT results to be the minimum concentration which achieved maximum inhibition within the range tested. For example, this was 3.7 μM for compound EM10 and 33.3 μM for EM100 (Figure 3*A*, *top two rows*).

The fluorometric SA-β-gal results are given in Figure 3*B*. In our hands, the fluorometric assay has log-linear relation with the proportion of HCT116 cells staining positive for SA-β-gal in the standard staining assay in the range up to ~25% positive cells at ~1.5-fold fluorescence induction (Supplementary Figure S6). We therefore imposed a cutoff of two-fold induction of signal relative to untreated cells. Fifteen compounds achieved greater than two-fold induction and were tested in colorimetric SA-β-gal staining dose responses.

The most potent effects were observed in both assays with EM100. SA-β-gal induction closely mirrored the inhibition of growth by MTT for this compound (MTT IC50, 9.6 μM; SA-β-gal EC50, 8.3 μM) (Figure 3*D*). EM100 is the ChemBridge compound 20903630 (Figure 3*C*, hereafter referred to as CB-20903630). Interestingly, at high concentration (33.3 μM), the compound was able to elicit detectable SA-β-gal staining in almost all cells against a background of extremely low (<1%) staining in untreated control cells (Figure 3*E*). Furthermore, staining in the entire population was achieved without observable loss of attachment, suggesting little or no cell death. Therefore, CB-20903630 appeared highly selective in inducing a key senescence marker.

Structure-Activity and Cell Cycle Inhibition Effects of Compound CB-20903630

To test preliminary SAR around CB-20903630, we searched commercial vendor libraries, identifying close structural analogues with a range of lipophilicities and functionalities, and obtained 10 structurally related analogues (Figure 4*A*). There were few commercially available analogues which maintained the 4-methyl-6-cyclobutyl motif present in CB-20903630, so we instead focused on analogues retaining the benzimidazolone motif.

Within this set (compounds 101-110), we found a range of MTT growth inhibition activities, with compound 101, containing the 1,2,4-triazine naphthyl group, possessing a respectable IC50 of 7.1 μM, in line with CB-20903630. Reassuringly, additional changes on this portion of the molecules were also tolerated, with both saturated (compounds 103 and 108) and unsaturated (compound 107) being tolerated. In addition, the presence of basic (compound 109) and neutral (compounds 102-106, 108, and 110) functionality indicates that there is potential to further optimize this series.

Because CB-20903630 remained among the best of the set tested, its identity and purity (>95%) were confirmed at resupply. We then investigated its effects on cell-cycle effectors involved in mediating arrest during senescence. HCT116 cells were treated for 48 hours in the presence of DMSO or 10 μM CB-20903630. Cells were harvested for Western blotting of cyclin B1, p21, CDK1, and CDC25C (Figure 4*B*). Cyclin B1, CDK1, and expression of the short isoform of CDC25C were all downregulated by CB-20903630 treatment, whereas levels of p21 were elevated (Figure 4*B*).

To confirm the cell cycle effects of the compound, we performed propidium iodide FACS analysis in treated HCT116 cells. Forty-eight–hour treatments at 20 μM were found to produce more robust effects than 10 μM, so we retained this dose for further growth-related assays. As shown in Figure 4*C*, CB-20903630 promoted a two-fold increase in G1 DNA content with a concomitant reduction in S-phase. Notably, there was no observed increase in the sub-G1 signal, suggesting that the compound does not significantly promote apoptosis and growth inhibition is primarily mediated through a G1 block.

We next compared growth inhibition of HCT116 with the isogenic deletion variants HCT116-p53$^{-/-}$ and HCT116-p21$^{-/-}$ (Figure 4*D*). The parental and the p53 deleted lines showed similar profiles, suggesting that p53 is not essential for cell cycle arrest by the compound. However, an approximately 1.5-fold reduction in sensitivity was observed in the p21 deleted cells. Hence, p21 but not p53 appears to play a role in the compound activity, in line with Figure 4*B*.

To confirm selectivity, we tested CB-20903630 in the M30 Apoptosense assay which measures an apoptotic neo-epitope of cleaved cytokeratin 18. CB-20903630 at 20 μM did not significantly increase cleaved CK18. However, the cytotoxic agent etoposide caused a 1.8-fold increase (Figure 4*E*, $P < .01$). Thus, CB-20903630 did not appear to induce substantial levels of apoptosis under these conditions in HCT116 cells.

Accelerated senescence is associated with an inflammatory phenotype characterized by secretion of a range of cytokines [6,7]. To investigate the inflammatory response of HCT116 cells, control or treated cell supernatants were tested in a multiplex assay analyzing the levels of 10 proinflammatory cytokines. Levels of most cytokines were low (Supplementary Figure S7) with the exception of IL8 (Figure 4*F*). CB-20903630 did not induce a proinflammatory

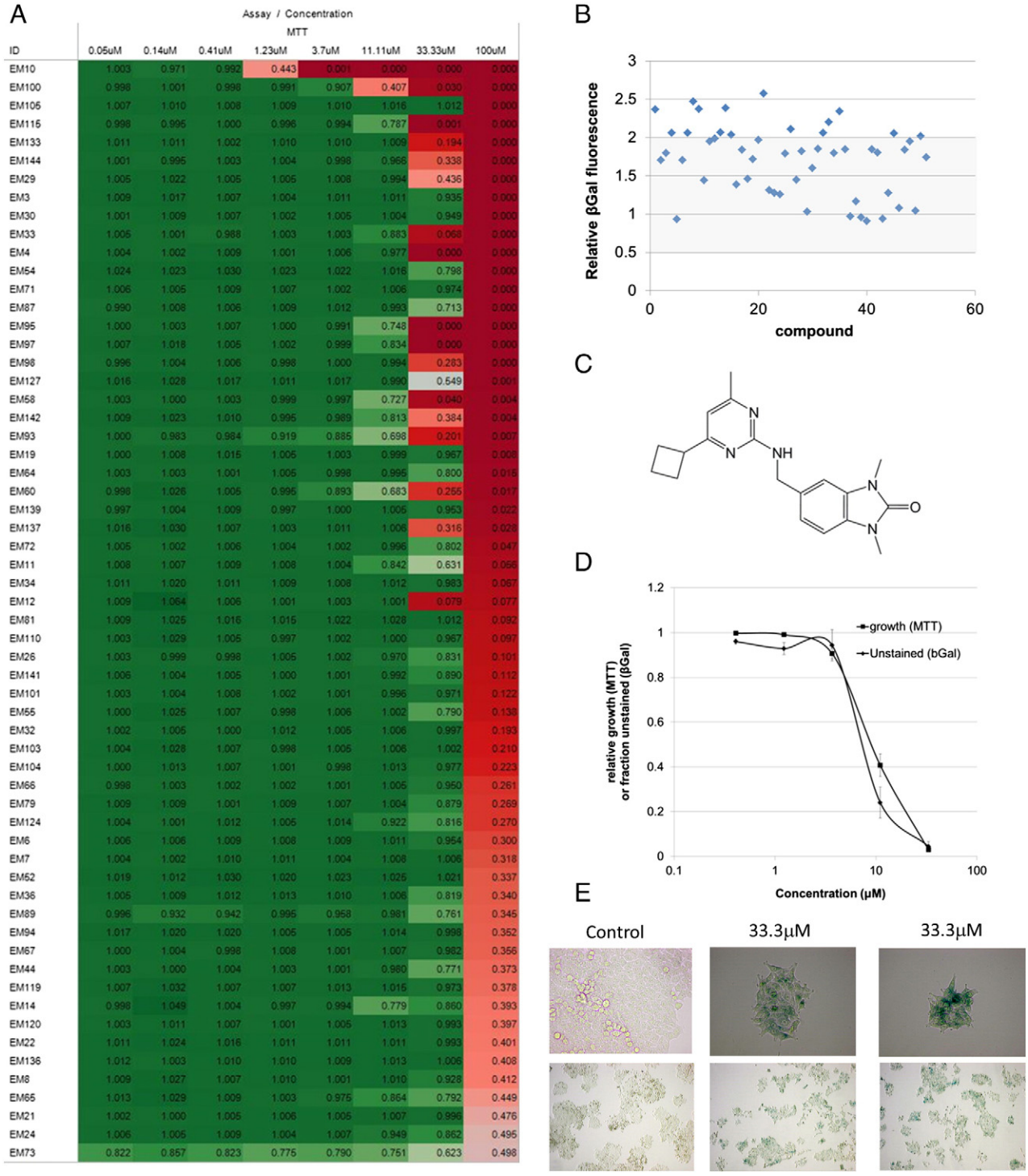


Figure 3. Cell-based screening results and identification of CB-20903630. (A) MTT cell growth inhibition results for compounds showing confirmed dose-dependent inhibition with IC50 < 100 μM. Heat maps were visualized in Tableau desktop. Numbers in each cell represent mean fold of control for each concentration of compound. Mean ± SEM of three experiments. (B) Fluorometric SA-β-gal on the 50 most potent MTT hits. Results are fold of vehicle-treated control. Two-fold activation of fluorescent signal was chosen as cutoff. Mean ± SEM of two experiments. (C) Structure of CB-20903630. The structure was confirmed by nuclear magnetic resonance, and purity was >95% by liquid chromatography. (D) Growth inhibition and SA-β-gal population-staining dose responses for CB-20903630. To clarify the shared dose response, data shown for SA-β-gal are unstained cells at each dose (1 minus SA-β-gal positive). Mean ± SEM of three experiments (MTT) or two experiments (SA-β-gal). (E) Representative micrographs showing SA-β-gal staining in untreated HCT116 cells or cells treated at 33.3 μM.

signature. Indeed, the only significant change was reduction in IL8 levels. Hence, despite its cell cycle effects, CB-20903630 did not induce the "senescence-associated secretory phenotype."

We next treated HCT116 cells continuously with DMSO or 20 μM CB-20903630 twice weekly for 1 month to determine cumulative population doublings with weekly counting (Figure 5*A*).

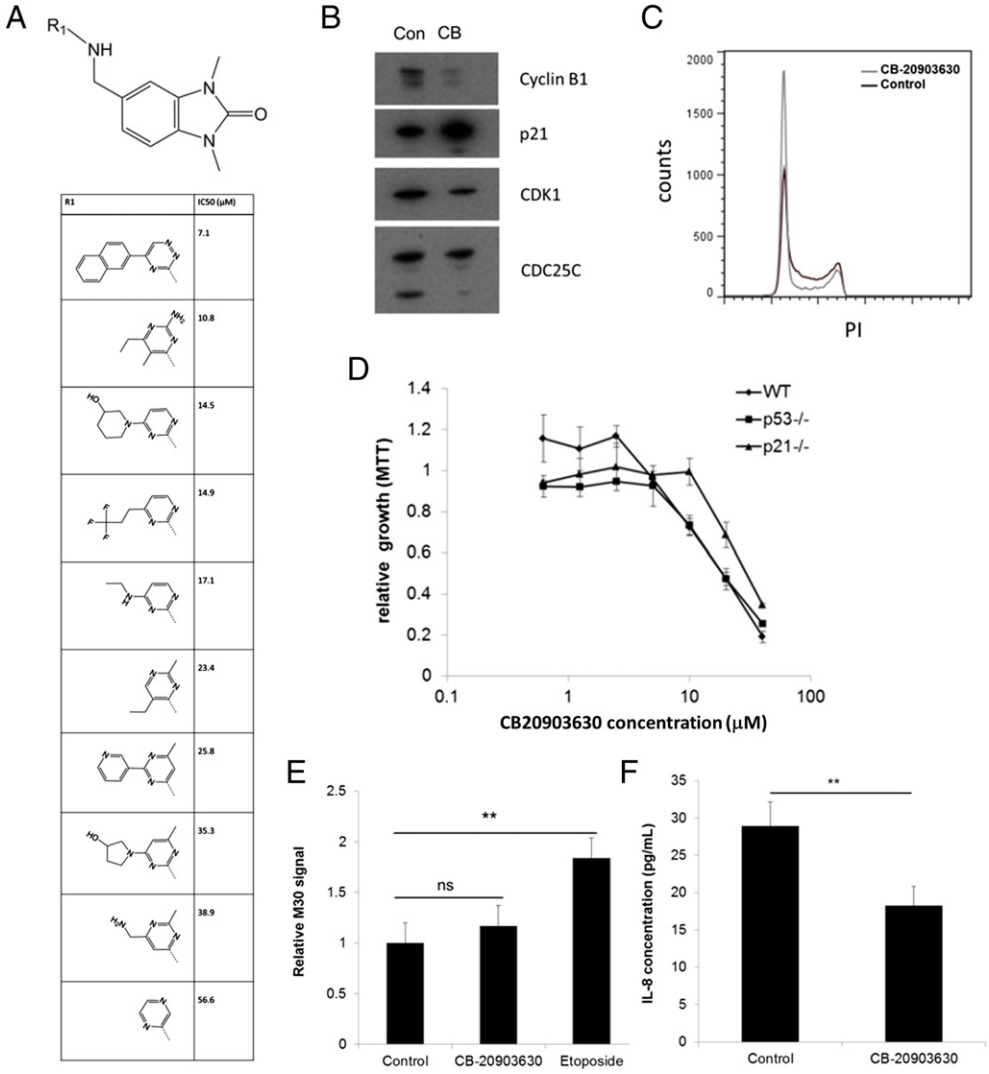


Figure 4. Structure-activity for the benzimidazolone scaffold and cell cycle effects of CB-20903630. (A) MTT SAR analysis of commercially available related analogues identified in the Chembridge catalogue. Mean IC50 of three experiments is shown. (B) Western blotting analysis of cell cycle effects in CB-20903630–treated HCT116 cells. Representative blots are shown. The experiment was performed twice. (C) Propidium iodide FACS analysis of cell cycle phase in control or treated cells. A representative histogram is shown. The experiment was performed three times. (D) MTT growth inhibition CB-20903630 dose-response in HCT116 or $p53^{-/-}$ and $p21^{-/-}$ isogenic variants. Mean ± SEM of three experiments. (E) Apoptosense CK18 assay of CB-20903630 or etoposide treatment in HCT116 cells. Mean ± SEM of three experiments (significance assessed by ANOVA: ns, not significant, $^{**}P < .01$). (F) Suppression of IL-8 levels by CB-20903630 in HCT116 cells. Mean ± SEM of three experiments (significance assessed by ANOVA: $^{**}P < .01$).

Control cells had undergone 33 population doublings by the end of treatment. In contrast, treated cell growth was severely retarded, and the cells underwent only 10.7 population doublings in total. Therefore, CB-20903630 treatment produced sustained inhibition of population growth.

We also examined CB-20903630 effects in an HCT116 spheroid model developed by adjustment to serum-free culture. Suspension cells were seeded for 5 days in the absence of treatment to allow initiation of multicellular spheroids, then swapped into 20 μM CB-20903630 or control medium (treatment day 0, Figure 5*B*), and cultured for a further 4 days. Treatment was repeated after 2 days. Compound was not removed between treatments.

Control spheroids significantly increased in volume in this period, whereas treated spheroids remained small (Figure 5*B*, *right panels*). Quantification of microscopic area of 50 individual spheroids in each condition showed a 3.5-fold difference (Figure 5*C*). Interestingly, following second treatment, many single cells were observed in treated flasks but not controls. In addition, treated spheroids were fewer in number and less tightly aggregated. It is possible that cell death pathways predominate under altered attachment.

Expression Profiling and Pathway Analysis of CB-20903630 Activity

To investigate the mechanism of action of CB-20903630, we performed microarray analysis using cDNA from cells treated for 48 hours with DMSO or 10 μM CB-20903630. We identified differentially expressed transcript IDs with greater than three-fold intensity change ($p < 0.05$) between control and treated cells. Modeling in MetaCore [28] generated a network of known direct interactions among the differentially expressed genes. All direct interactions with cluster size ≥ 2 were included (Figure 6*A*).

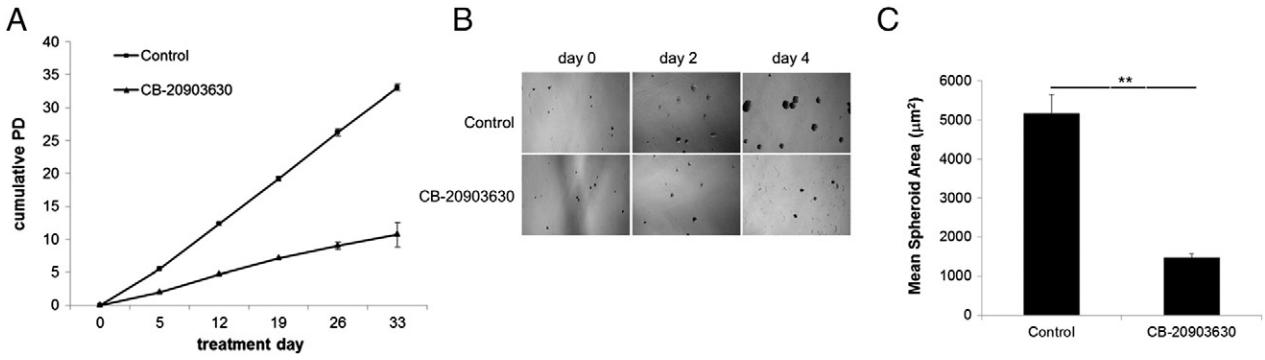


Figure 5. Long-term growth effects of repeat treatment with CB-20903630 and inhibition of multicellular spheroid growth. (A) HCT116 cells were maintained in culture and treated twice-weekly with CB-20903630 or vehicle. Cell numbers were counted weekly for calculation of cumulative population doublings. Mean ± SEM of three experiments. (B) HCT116 cells were adapted to serum-free conditions to generate a suspension line which grows as multicellular spheroids. Small spheroids were allowed to form in culture medium for 5 days then treated twice with CB-20903630 or vehicle. Representative micrographs obtained during the treatment period are shown. (C) Quantitation of mean area of 50 treated or control spheroids after 4 days of treatment with CB-20903630. Significance of population difference was assessed by Wilcoxon rank sum test (**$P < .001$).

Several transcription factors associated with development and proliferation were affected, including upregulation of Fra-2, DBP, and C/EBP, whereas E2F2, Gli-1, MEIS2, PBX, and Sox4 were downregulated. Interestingly, a number of secreted and membrane proteins were also downregulated including CCL19, vasohibin 2, semaphorins, fibulin-5, MMP9, BMP4, and ephrin A. These results suggest that the compound may regulate a secretory program distinct from the inflammatory markers investigated above. Clock genes PER1 and PER3 were also differentially regulated, in line with a previous study which found clock gene repression in vascular smooth muscle cells undergoing telomere-dependent senescence [36]. Process enrichment indicated that CB-20903630 promotes differential expression on inflammatory and developmental signaling networks as suggested by the model (Figure 6*B*).

CB-20903630 contains a kinase hinge-binding motif [37], indicating that the compound may target a cell cycle–related kinase. We generated expression profiles of IMR90 fibroblasts treated with 13 well-characterized kinase inhibitors (Supplementary Table S3) most of which induce a senescence-associated heterochromatin foci–like phenotype [49] and apoptosis and/or cell cycle responses. This data set represents a range of senescence effects induced by different pathway-specific agents in cells with an intact senescence response.

To compare the effects of CB-20903630, we also treated IMR90 with the compound and compared both HCT116 and IMR90 profiles with the other inhibitors. Responses were clustered on significance of overlap in affected MetaCore process networks. Numbers of significant process networks under each inhibitor treatment were used to generate hypergeometric probabilities for each pairwise comparison, which we used as an unweighted average distance metric (Figure 6*C*).

EGFR inhibitor clustered with JNKIX. Cell cycle pathways, cell adhesion, and developmental and cytoskeletal processes were affected by this group. Two AKT inhibitors (AKTV and AKTVIII) are present in the analysis, alongside two PI3K inhibitors (PI103 and GDC0941). AKTV/GDC0941 clustered and the CB-20903630 process network profile in IMR90 cells also clustered in this group. Adhesion, inflammation, development, and proteolysis processes are strongly represented in this group. The AKTVIII profile was also close to these in the analysis, whereas the other PI3K inhibitor PI103 was more closely related to the PDGFR inhibitor; both of these had very large process network profiles (52 and 59 networks affected, respectively). MAPK inhibitor MK2A clustered with Src-family inhibitor SU6656 and with AuroraII, primarily affecting DNA damage, cell cycle, and apoptosis processes.

The CB-20903630 process network profile in HCT116 clustered away from all others, possibly indicating that the compound mechanism affects different pathways in normal versus cancer cells. The profile in IMR90 appears to suggest similarity with agents targeting the PI3K/AKT pathway. CB-20903630 profiles in IMR90 and HCT116 were partially overlapping because five of the eight processes scoring as significant in HCT116 cells were also significant in IMR90 cells (Figure 6, *A* and *B*, and Supplementary Figures S8 and S9). Inflammatory processes were also highly represented in IMR90. However, IMR90 cells also scored highly in a range of development and proteolysis processes shared by the PI3K/AKT agents, making the observed profile more similar to these. Thus, different pathways may be affected by the compound in normal versus cancer cells.

To determine whether CB-20903630 is structurally related to existing kinase inhibitors, we performed clustering analysis on 3D pharmacophores comparing CB-20903630 alongside 527 known kinase inhibitors using a self-organizing map (Figure 6*D*). CB-20903630 loaded with 15 other compounds on a neuron which did not cluster strongly with neighbors. Examination of the structures showed prevalence of JNK2/3, VEGFR2, and GSK3 inhibitors (Supplementary File 2). However, CB-20903630 had little 2D similarity with these. Together, our results suggest that CB-20903630 is a selective cell cycle inhibitor which appears to be structurally novel.

Discussion

Cellular immortality is a hallmark of cancer and a near-universal cancer target. However, recent clinical results suggest that telomerase may prove a more refractory target than had been hoped in solid tumors [38]. Multiple pathways regulate telomerase, and a variety of backup mechanisms may exist facilitating escape from inhibition [28,39]. On the other hand, strong interest in senescence induction as an alternative target to reverse limitless replicative potential of cancer cells has also emerged in recent years.

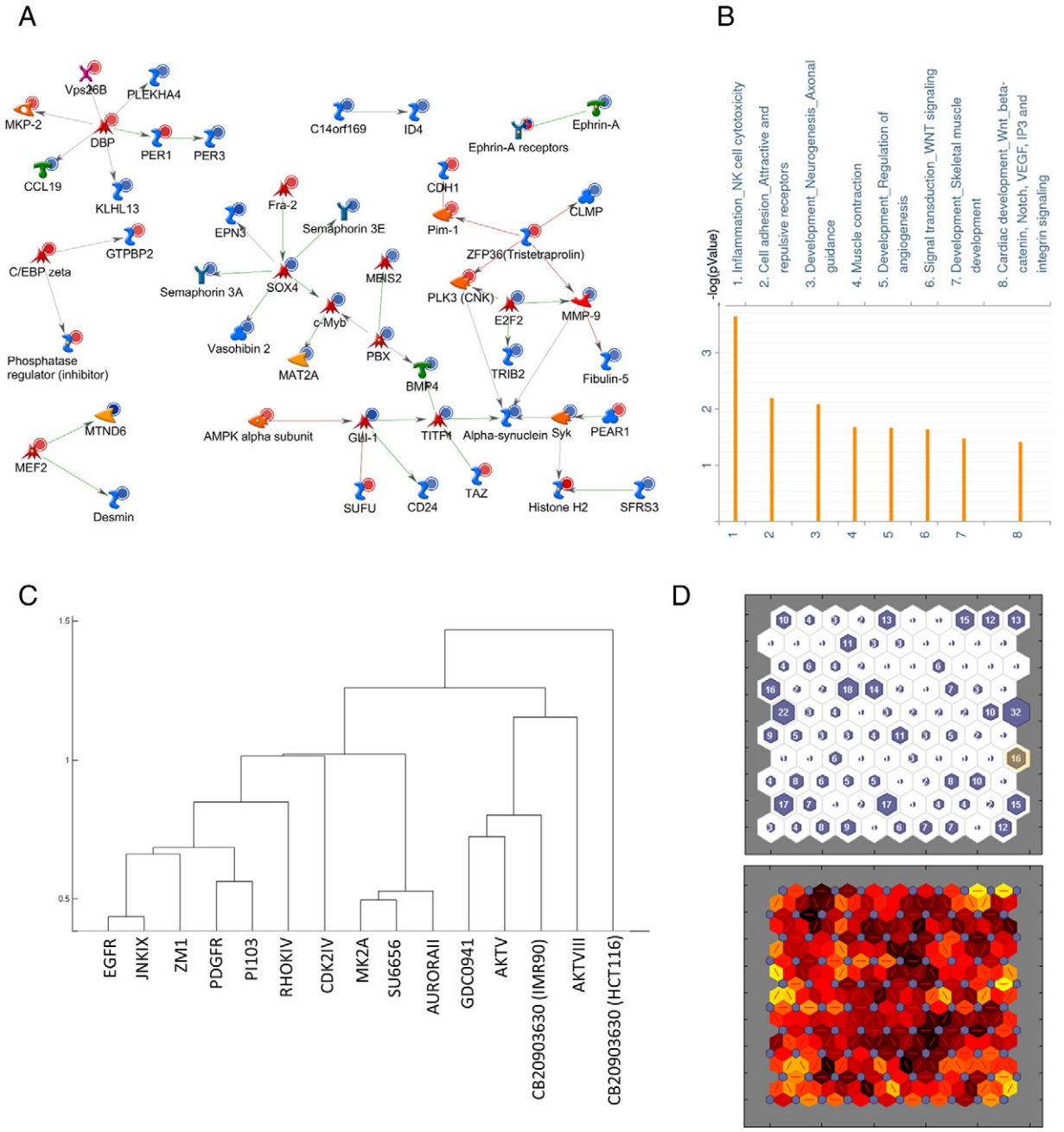


Figure 6. Microarray and structural analysis of CB-20903630. (A) Direct interactions network of differentially expressed genes in HCT116 cells treated with 10 μM CB-20903630. RNA samples from DMSO versus compound-treated cells were profiled on Agilent whole genome expression arrays. Differentially expressed gene lists were analyzed in MetaCore by the direct interactions algorithm to obtain the network model. Green and red arrows indicate known activating or inhibitory interactions between entities, respectively. Red and blue circles indicate upregulation and downregulation of expression relative to vehicle treatment, respectively. (B) Significant differentially affected GeneGo process networks under CB-20903630 treatment in HCT116 cells obtained by enrichment analysis of differentially affected genes. (C) Clustering of process network profiles with cumulative hypergeometric probability of pairwise overlap as the unweighted distance metric. (D) SOM structural clustering of CB-20903630 and known kinase inhibitors (see Materials and Methods section for source of comparator structures) after 200 training cycles. (Upper panel) Compound loadings: CB-20903630 and 15 other compounds loaded on the highlighted neuron in the upper panel. Numbers indicate number of compounds on each neuron. (Lower panel) Visualization of neighbor weights: the CB-20903630 neuron is not strongly clustered with its neighbors (darker bands indicate larger distances).

We began with a strategy to match training compound sets to the expression profile of telomere dysfunction. We identified a target panel, optimized a neural network ensemble, and screened a 2M-compound virtual library. Virtual hits were prioritized based on ADMET filters and pharmacophore clustering to identify a cell-based screening set, resulting in identification of

CB-20903630. The compound promotes SA-β-gal in the majority of HCT116 cells and modulates cell cycle targets p21, cyclin B1, CDC25C, and CDK1, causing a G1 blockade without observable induction of cell death.

Preliminary SAR indicates scope for enhancement of CB-20903630 activity. However, relatively few close commercial analogues were found to be available, and we have therefore not exhaustively investigated this aspect and focused instead on the activity of CB-20903630. The growth inhibition effect appears to be in part dependent on p21 but not p53 based on sensitivities of isogenic HCT116 variants. In long-term treatments, the compound severely repressed population doubling times and strongly repressed growth in a multicellular spheroid model. The target of CB-20903630 is unknown. However, expression and structural clustering analysis suggest that the effects of the compound in normal cells have similarity with agents targeting the PI3K/AKT pathway, whereas in cancer cells, the effects diverged from other well characterized cell cycle inhibitors.

Virtual screening has previously been used for the identification of ligands for single targets [40]. In one recent example, the compound 6,6"-biapigenin was identified as a novel inhibitor of the NEDD8 activating enzyme (NAE) which is required for NEDDylation of a wide range of cellular targets. A previously identified NAE inhibitor showed broad activity against cancer cell lines, and the authors used molecular docking against a quaternary complex comprising the NAE subunits APPBP1 and UBA3 as well as its NEDD8 and ATP substrates to identify the new compound which showed low micromolar activity in Caco-2 cells [41]. Multitarget approaches such as that described here have not been widely investigated, although a recent study reported use of combinatorial support vector machine classifiers to identify dual-specificity ligands for a range of kinase pairs [42]. Furthermore, "inverse docking" in which individual compounds are docked against target panels has been suggested as a potentially powerful tool for compound repositioning strategies to complement existing pipelines in the pharmaceutical industry [43]. Our results suggest that virtual screening focused on target panels may also provide a useful approach for the identification of phenotype-focused libraries.

Ultimately, development of senescence therapeutics will require greater understanding of the regulation of senescence signaling networks. CB-20903630 is an interesting tool compound which appears to be a highly selective cell cycle inhibitor. The compound may therefore be a useful probe to identify new candidate markers and mechanisms associated with senescence and cell cycle responses. However, broad knowledge of the scaffold types that are able to regulate senescence pathways will also be required to identify a range of selective agents. In this paper, we identified a kinase-like scaffold by restricting our training set and library to "drug-like" chemical space. However, other regions of chemical space might also be worthy of consideration, such as natural product libraries or novel organometallic agents [44,45]. As in the current report, virtual screening might also provide an extremely useful tool to probe these novel library types in future studies.

Materials and Methods

Training Library Assembly and Neural Network Optimization

Training compounds were identified in confirmatory PubChem bioassay entries reported in Table 1 or from in-house data in the case of p21. All neural network optimization on neuron number, number of hidden layers, and training rule was performed using the Matlab neural network toolbox (Mathworks, Natick, MA). Ten-fold cross-validation was performed on each parameter set. The overall optimization workflow is shown in Supplementary Figure S2. Chemical descriptor sets were obtained using PaDEL Descriptor [33] and PowerMV [32].

Highly correlated descriptors and those with a variance of 0 were excluded, and the smaller representative set of descriptors was chosen using the Feature Selection option of the Canvas program [46,47] (Canvas, version 2.0; Schrödinger, LLC, New York, NY). 3D Pharmacophore fingerprints were also calculated in Canvas and ADME/Tox properties (Supplementary Table S2) in QikProp (QikProp, version 4.0; Schrödinger, LLC, New York, NY, 2014).

Cell Lines and Compounds

The cells used were HCT116 colorectal cancer cells, their p53 and p21 deleted isogenic derivatives, and IMR90 fibroblasts. One hundred forty-seven virtual hits were selected based on the clustering analysis using 3D pharmacophore fingerprints. Compounds were initially sourced through E-Molecules (Stevenage, UK). CB-20903630 resupply and related analogues were obtained from ChemBridge (San Diego, CA). Structure was confirmed by nuclear magnetic resonance, and purity was confirmed by liquid chromatography/mass spectrometry. All other signal transduction inhibitors reported were obtained from Millipore (Supplementary Table S3). GRN163L was kindly provided by the Geron Corporation (Menlo Park, CA).

MTT, Fluorescent SA-β-gal, and M30 Assays

For MTT assay, cells were treated twice over 48 hours with compounds and then incubated for a further 3 days before MTT assay (MTT supplied by Sigma, Dorset, UK). MTT reduction assays were performed using Softmax Pro software (Molecular Devices Ltd., Wokingham, UK). All experiments were repeated three times. Heat maps were generated in Tableau Desktop (Tableau Software, Seattle, WA). Fluorescent SA-β-gal assays were performed using the 96-well kit by Cell Biolabs (San Diego, CA). Cells were seeded overnight before 48-hour compound treatments. Five micrograms of protein was incubated in duplicate with assay buffer for 3 hours. Fluorescence was measured using a Safire plate reader (Tecan Group, Männedorf, Switzerland). All experiments were repeated twice. In colorimetric SA-β-gal staining assays, cells were fixed in gluteraldehyde and stained in the dark overnight with X-gal at pH 6. At least 500 cells in 5 random fields were counted for microscopic evaluation of proportions of stained cells in any treatment condition.

For M30 assay, cells were seeded in triplicate wells of 96-well plates overnight before addition of compound. Cells were treated for 2 days with compounds or vehicle then harvested. ELISA was performed on supernatants according the manufacturer's instructions (VLVbio, Nacka, Sweden) with quantification using Softmax Pro software (Molecular Devices Ltd., Wokingham, UK). Experiments were repeated three times.

Western Blotting

Twenty micrograms of protein was separated by SDS-PAGE, blotted onto polyvinylidene difluoride (Millipore, Watford, UK), and blocked overnight in PBS-T containing 5% nonfat dried milk. Antibodies were cyclin B1 (4135), p21 (2946), CDK1 (9112), and

CDC25C (4688), all obtained from New England Biolabs UK (Hitchin, UK). Primary antibodies were detected with HRP-conjugated secondary. HRP was detected using ECL detection reagents (Amersham Pharmacia, Buckinghamshire, UK). Experiments were performed at least twice.

FACS Analysis

Treated cells were fixed in 70% ethanol and stained with 0.05 mg/ml of propidium iodide solution containing 1 mg/ml of RNAse A (both obtained from Life Technologies, Paisley, UK) and 0.3% Tween-20 (Sigma, Dorset, UK). FACS was performed on a FACSverse instrument (BD Biosciences, Oxford, UK). Experiments were performed three times.

Microarray Processing

RNA was labeled and amplified using the one-color microarray gene expression analysis protocol (Agilent Technologies, Santa Clara, CA), hybridized to Agilent whole human genome 4 × 44k Agilent whole human genome microarrays, and incubated for 17 hours at 60°C in a hybridization oven. Arrays were washed on a magnetic stirrer using Agilent wash buffers. Slides were scanned on an Agilent microarray scanner at 5-μm resolution, photomultiplier tube (PMT) gain at 100% and 10%. The extended dynamic range setting was corrected for saturation. Kinase inhibitor treatments in IMR90 were performed twice. CB-20903630 treatment of HCT116 was performed three times, but in IMR90 cells, only two independent experiments were performed because of compound availability. The data set has been submitted to the Gene Expression Omnibus with accession number GSE72621.

Microarray Data Analysis

Microarray data were extracted using Agilent Feature Extraction software (Agilent Technologies, Santa Clara, CA). All array data were analyzed in GeneSpring for normalization and statistical analysis (Agilent Technologies, Santa Clara, CA). Intraarray normalization was carried out using the 75th percentile for each microarray. Significant differences in expression between control and treated cells were determined using unpaired t test. IDs with greater than three-fold intensity change, $P < .05$, were selected for further analysis.

Process Profiles, Network Modeling, and Structural Clustering

Differentially expressed genes were analyzed in MetaCore (Thomson Reuters, New York, NY) using enrichment analysis by GeneGo process networks [27]. Probability of overlap was scored for each gene list tested against all process networks in the MetaCore database. Network analyses were performed using the GeneGo direct interactions algorithm [29]. For process network clustering, cumulative hypergeometric probability of pairwise overlap between process network profiles was calculated. Dendrograms were generated from the matrix of pairwise probabilities for all comparisons using unweighted average distances in Matlab (Mathworks, Natick, MA).

For structural clustering, 3D pharmacophores were calculated in Canvas. Included structures were CB-20903630 and 527 kinase inhibitors pooled from the Millipore Inhibitor Select 384-well panel (#539743) and from the GSK Published Kinase Inhibitor Set [48]. The finger prints were clustered on a 10 × 10 self-organizing map in Matlab using 200 training iterations. Compounds loading with CB-20903630 are given in Supplementary File 2.

Statistical Analysis

All statistical analyses were performed in Microsoft Excel or Matlab.

Acknowledgements

This work was supported by a Development Fund award from the Cancer Research UK Glasgow Centre, the Glasgow Experimental Cancer Medicine Centre (funded by Cancer Research UK and the Chief Scientist Office, Scotland), Cancer Research UK grants C2193/A15584 and C301/A12962, University of Cambridge, Cancer Research UK Cambridge Institute Core Grant, and Hutchison Whampoa (M.N.). The authors would like to thank the Geron Corporation (Menlo Park, CA) for their kind gift of GRN163L.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neo.2015.08.009.

References

[1] Cairney CJ, Bilsland AE, Evans TR, Roffey J, Bennett DC, Narita M, Torrance CJ, and Keith WN (2012). Cancer cell senescence: a new frontier in drug development. *Drug Discov Today* **17**, 269–276.

[2] Bilsland AE, Cairney CJ, and Keith WN (2011). Targeting the telomere and shelterin complex for cancer therapy: current views and future perspectives. *J Cell Mol Med* **15**, 179–186.

[3] Lafferty-Whyte K, Bilsland A, Hoare SF, Burns S, Zaffaroni N, Cairney CJ, and Keith WN (2010). TCEAL7 inhibition of c-Myc activity in alternative lengthening of telomeres regulates hTERT expression. *Neoplasia* **12**, 405–414.

[4] Nardella C, Clohessy JG, Alimonti A, and Pandolfi PP (2011). Pro-senescence therapy for cancer treatment. *Nat Rev Cancer* **11**, 503–511.

[5] Lafferty-Whyte K, Bilsland A, Cairney CJ, Hanley L, Jamieson NB, Zaffaroni N, Oien KA, Burns S, Roffey J, and Boyd SM, et al (2010). Scoring of senescence signalling in multiple human tumour gene expression datasets, identification of a correlation between senescence score and drug toxicity in the NCI60 panel and a pro-inflammatory signature correlating with survival advantage in peritoneal mesothelioma. *BMC Genomics* **11**, 532.

[6] Bilsland AE, Revie J, and Keith W (2013). MicroRNa and senescence: the senectome, integration and distributed control. *Crit Rev Oncog* **18**, 373–390.

[7] Pawlikowski JS, Adams PD, and Nelson DM (2013). Senescence at a glance. *J Cell Sci* **126**, 4061–4067.

[8] Ohtani N, Takahashi A, Mann DJ, and Hara E (2012). Cellular senescence: a double-edged sword in the fight against cancer. *Exp Dermatol* **21**(Suppl. 1), 1–4.

[9] Zhang Y and Yang JM (2011). The impact of cellular senescence in cancer therapy: is it true or not? *Acta Pharmacol Sin* **32**, 1199–1207.

[10] Bitler BG, Fink LS, Wei Z, Peterson JR, and Zhang R (2013). A high-content screening assay for small-molecule modulators of oncogene-induced senescence. *J Biomol Screen* **18**, 1054–1061.

[11] Lahtela J, Corson LB, Hemmes A, Brauer MJ, Koopal S, Lee J, Hunsaker TL, Jackson PK, and Verschuren EW (2013). A high-content cellular senescence screen identifies candidate tumor suppressors, including EPHA3. *Cell Cycle* **12**, 625–634.

[12] Rovillain E, Mansfield L, Lord CJ, Ashworth A, and Jat PS (2011). An RNA interference screen for identifying downstream effectors of the p53 and pRB tumour suppressor pathways involved in senescence. *BMC Genomics* **12**, 355.

[13] Ewald JA, Desotelle JA, Wilding G, and Jarrard DF (2010). Therapy-induced senescence in cancer. *J Natl Cancer Inst* **102**, 1536–1546.

[14] Ewald JA and Jarrard DF (2012). Decreased skp2 expression is necessary but not sufficient for therapy-induced senescence in prostate cancer. *Transl Oncol* **5**, 278–287.

[15] Sudha A, Srinivasan P, and Rameshthangam P (2014). Exploration of potential EGFR inhibitors: a combination of pharmacophore-based virtual screening, atom-based 3D-QSAR and molecular docking analysis. *J Recept Signal Transduct Res* , 1–12.

[16] Zakharov AV, Peach ML, Sitzmann M, Filippov IV, McCartney HJ, Smith LH, Pugliese A, and Nicklaus MC (2012). Computational tools and resources for metabolism-related property predictions. 2. Application to prediction of half-life time in human liver microsomes. *Future Med Chem* **4**, 1933–1944.

[17] Liao C, Sitzmann M, Pugliese A, and Nicklaus MC (2011). Software and resources for computational medicinal chemistry. *Future Med Chem* **3**, 1057–1085.

[18] Mueller R, Dawson ES, Meiler J, Rodriguez AL, Chauder BA, Bates BS, Felts AS, Lamb JP, Menon UN, and Jadhav SB, et al (2012). Discovery of 2-(2-benzoxazoyl amino)-4-aryl-5-cyanopyrimidine as negative allosteric modulators (NAMs) of metabotropic glutamate receptor 5 (mGlu(5)): from an artificial neural network virtual screen to an in vivo tool compound. *ChemMedChem* **7**, 406–414.

[19] Geppert H, Vogt M, and Bajorath J (2010). Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation. *J Chem Inf Model* **50**, 205–216.

[20] Afantitis A, Melagraki G, Koutentis PA, Sarimveis H, and Kollias G (2011). Ligand-based virtual screening procedure for the prediction and the identification of novel beta-amyloid aggregation inhibitors using Kohonen maps and Counterpropagation Artificial Neural Networks. *Eur J Med Chem* **46**, 497–508.

[21] Jorissen RN and Gilson MK (2005). Virtual screening of molecular databases using a support vector machine. *J Chem Inf Model* **45**, 549–561.

[22] Warmuth MK, Liao J, Ratsch G, Mathieson M, Putta S, and Lemmen C (2003). Active learning with support vector machines in the drug discovery process. *J Chem Inf Comput Sci* **43**, 667–673.

[23] Moller MF (1993). A scaled conjugate-gradient algorithm for fast supervised learning. *Neural Netw* **6**, 525–533.

[24] Wang Y, Bolton E, Dracheva S, Karapetyan K, Shoemaker BA, Suzek TOWang J, Xiao J, Zhang J, and Bryant SH (2010). An overview of the PubChem BioAssay resource. *Nucleic Acids Res* **38**, D255–D266.

[25] Wang Y, Suzek T, Zhang J, Wang J, He S, Cheng T, Shoemaker BA, Gindulyte A, and Bryant SH (2014). PubChem BioAssay: 2014 update. *Nucleic Acids Res* **42**, D1075–D1082.

[26] Kim MM, Rivera MA, Botchkina IL, Shalaby R, Thor AD, and Blackburn EH (2001). A low threshold level of expression of mutant-template telomerase RNA inhibits human tumor cell proliferation. *Proc Natl Acad Sci U S A* **98**, 7982–7987.

[27] Bessarabova M, Ishkin A, JeBailey L, Nikolskaya T, and Nikolsky Y (2012). Knowledge-based analysis of proteomics data. *BMC Bioinformatics* , 13.

[28] Bilsland AE, Hoare S, Stevenson K, Plumb J, Gomez-Roman N, Cairney C, Burns S, Lafferty-Whyte K, Roffey J, and Hammonds T, et al (2009). Dynamic telomerase gene suppression via network effects of GSK3 inhibition. *PLoS One* **4**, e6459.

[29] Ekins S, Bugrim A, Brovold L, Kirillov E, Nikolsky Y, Rakhmatulin E, Sorokina S, Ryabov A, Serebryiskaya T, and Melnikov A, et al (2006). Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. *Xenobiotica* **36**, 877–901.

[30] Gomez-Millan J, Goldblatt EM, Gryaznov SM, Mendonca MS, and Herbert BS (2007). Specific telomere dysfunction induced by GRN163L increases radiation sensitivity in breast cancer cells. *Int J Radiat Oncol Biol Phys* **67**, 897–905.

[31] Herbert BS, Gellert GC, Hochreiter A, Pongracz K, Wright WE, Zielinska D, Chin AC, Harley CB, Shay JW, and Gryaznov SM (2005). Lipid modification of GRN163, an N3'–>P5' thio-phosphoramidate oligonucleotide, enhances the potency of telomerase inhibition. *Oncogene* **24**, 5262–5268.

[32] Liu K, Feng J, and Young SS (2005). PowerMV: a software environment for molecular viewing, descriptor generation, data analysis and hit evaluation. *J Chem Inf Model* **45**, 515–522.

[33] Yap CW (2011). PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. *J Comput Chem* **32**, 1466–1474.

[34] Carletta J (1996). Assessing agreement on classification tasks: the kappa statistic. *Comput Linguist* **22**, 249–254.

[35] Landis JR and Koch GG (1977). The measurement of observer agreement for categorical data. *Biometrics* **33**, 159–174.

[36] Kunieda T, Minamino T, Katsuno T, Tateno K, Nishi J, Miyauchi H, Orimo M, Okada S, and Komuro I (2006). Cellular senescence impairs circadian expression of clock genes in vitro and in vivo. *Circ Res* **98**, 532–539.

[37] Liao JJ (2007). Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J Med Chem* **50**, 409–424.

[38] Chiappori A, Kolevska T, Burington B, Spigel D, Hager S, Rarick M, Gadgeel S, Blais N, Von Pawel J, and Hart L, et al (2013). A randomized phase II study of the telomerase inhibitor imetelstat as maintenance therapy for advanced non-small cell lung cancer. *Cancer Res* **73**.

[39] Bilsland AE, Stevenson K, Liu Y, Hoare S, Cairney CJ, Roffey J, and Keith WN (2014). Mathematical model of a telomerase transcriptional regulatory network developed by cell-based screening: analysis of inhibitor effects and telomerase expression mechanisms. *PLoS Comput Biol* **10**, e1003448.

[40] Clark DE (2008). What has virtual screening ever done for drug discovery? *Expert Opin Drug Discov* **3**, 841–851.

[41] Leung CH, Chan DS, Yang H, Abagyan R, Lee SM, Zhu GY, Fong WF, and Ma DL (2011). A natural product-like inhibitor of NEDD8-activating enzyme. *Chem Commun* **47**, 2511–2513.

[42] Ma XH, Wang R, Tan CY, Jiang YY, Lu T, Rao HB, Li XY, Go ML, Low BC, and Chen YZ (2010). Virtual screening of selective multitarget kinase inhibitors by combinatorial support vector machines. *Mol Pharm* **7**, 1545–1560.

[43] Ma DL, Chan DS, and Leung CH (2013). Drug repositioning by structure-based virtual screening. *Chem Soc Rev* **42**, 2130–2141.

[44] Leung CH, Zhong HJ, Chan DS, and Ma DL (2013). Bioactive iridium and rhodium complexes as therapeutic agents. *Coord Chem Rev* **257**, 17644–17676.

[45] Ma DL, Chan DS, and Leung CH (2011). Molecular docking for virtual screening of natural product databases. *Chem Sci* **2**, 1656–1664.

[46] Duan J, Dixon SL, Lowrie JF, and Sherman W (2010). Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J Mol Graph Model* **29**, 157–170.

[47] Sastry M, Lowrie JF, Dixon SL, and Sherman W (2010). Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *J Chem Inf Model* **50**, 771–784.

[48] Knapp S, Arruda P, Blagg J, Burley S, Drewry DH, Edwards A, Fabbro D, Gillespie P, Gray NS, and Kuster B, et al (2013). A public-private partnership to unlock the untargeted kinome. *Nat Chem Biol* **9**, 3–6.

[49] Sadaie M, Dillon C, Narita M, Young AR, Cairney CJ, Godwin LS, Torrance CJ, Bennett DC, Keith WN, and Narita M (2015). Cell-based screen for altered nuclear phenotypes reveals senescence progression in polyploid cells after Aurora kinase B inhibition. *Mol Biol Cell* **26**, 2971–2985.

NEOPLASIA
www.neoplasia.com

Volume 16 Number 7 July 2014 pp. 606–615 606

Immortalization of T-Cells Is Accompanied by Gradual Changes in CpG Methylation Resulting in a Profile Resembling a Subset of T-Cell Leukemias[1,2,3]

Sofie Degerman*,[4], Mattias Landfors*,[¶,4], Jan Konrad Siwicki[†], John Revie[§], Magnus Borssén*, Emma Evelönn*, Erik Forestier*, Krystyna H. Chrzanowska[‡], Patrik Rydén[¶,#], W. Nicol Keith[§] and Göran Roos*

*Department of Medical Biosciences, Umeå University, SE-90185 Umeå, Sweden; [†]Department of Immunology, Maria Sklodowska-Curie Memorial Cancer Centre and Institute of Oncology, 02-781 Warsaw, Poland; [‡]Department of Medical Genetics, Children's Memorial Health Institute, 04-730 Warsaw, Poland; [§]Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Glasgow G61 1QH, UK; [¶]Department of Mathematics and Mathematical Statistics, Umeå University, SE-90185 Umeå, Sweden; [#]Computational Life Science Cluster, Umeå University, SE-90185 Umeå, Sweden

Abstract

We have previously described gene expression changes during spontaneous immortalization of T-cells, thereby identifying cellular processes important for cell growth crisis escape and unlimited proliferation. Here, we analyze the same model to investigate the role of genome-wide methylation in the immortalization process at different time points pre-crisis and post-crisis using high-resolution arrays. We show that over time in culture there is an overall accumulation of methylation alterations, with preferential increased methylation close to transcription start sites (TSSs), islands, and shore regions. Methylation and gene expression alterations did not correlate for the majority of genes, but for the fraction that correlated, gain of methylation close to TSS was associated with decreased gene expression. Interestingly, the pattern of CpG site methylation observed in immortal T-cell cultures was similar to clinical T-cell acute lymphoblastic leukemia (T-ALL) samples classified as CpG island methylator phenotype positive. These sites were highly overrepresented by polycomb target genes and involved in developmental, cell adhesion, and cell signaling processes. The presence of non-random methylation events in *in vitro* immortalized T-cell cultures and diagnostic T-ALL samples indicates altered methylation of CpG sites with a possible role in malignant hematopoiesis.

Neoplasia (2014) 16, 606–615

Introduction

Cellular immortalization is a multistep process and a major step in cancer development. Senescence checkpoint bypass and acquisition of indefinite replicative capacity in cell cultures have been associated with pathways affecting cell cycle progression, DNA damage, oxidative stress responses, and cytoskeletal organization, as well as

Abbreviations: CIMP, CpG island methylator phenotype; DEG, differently expressed gene; DM-CpG, differently methylated CpG; DMG, differently methylated gene; T-ALL, T-cell acute lymphoblastic leukemia

Address all correspondence to: Sofie Degerman, PhD, Department of Medical Biosciences, Umeå University, SE-90185 Umeå, Sweden. E-mail: sofie.degerman@medbio.umu.se

[1] This article refers to supplementary materials, which are designated by Supplementary Figure S1 to S5 and are available online at http://www.neoplasia.com.

[2] Author contributions: S.D., M.L., W.N.K., and G.R. conceived and designed the experiments. S.D., J.K.S., E.E., and K.H.C. performed the experiments. S.D., M.L., P.R., E.F., M.B., and J.R. analyzed the data. S.D., M.L., J.K.S., W.N.K., and G.R., with contribution from coauthors, wrote the paper.

[3] Conflict-of-interest disclosure: The authors declare no competing financial interests.

[4] These authors contributed equally.

Received 16 May 2014; Revised 1 July 2014; Accepted 3 July 2014

1476-5586/14
http://dx.doi.org/10.1016/j.neo.2014.07.001

interferon-, insulin growth factor–, and MAP kinase–related pathways [1]. Published data on cellular processes involved in immortalization have essentially been generated from fibroblasts and endothelial and epithelial cells, whereas less is known for lymphoid cells [1–4]. A common feature of immortalized cells is activation of telomerase through up-regulation of human Telomerase Reverse Transcriptase (hTERT), the expression of which is regulated by a multitude of factors including modifiers of the chromatin structure [5–7]. We have previously shown that impaired DNA damage response and deregulated cell senescence control together with activation of telomerase were coupled to T-cell immortalization [8]. *In vitro* immortalization of mammary epithelial cells has been associated with stepwise DNA methylation alterations [3], and in the present study, we have analyzed methylation alterations during this process in primary T-cell cultures and in relation to diagnostic T-cell acute lymphoblastic leukemia (T-ALL) samples.

Epigenetic processes involve DNA methylation and histone modifications, which can participate in gene regulation without altering the DNA sequence. DNA methylation frequently occurs on a cytosine followed by a guanine (CpG sites) [9]. Many CpG-enriched regions (CpG islands) are located in promoters and methylation of such CpG islands represents one major transcriptional control mechanism [4]. Abnormal DNA methylation is a hallmark of cancer development and might lead to silencing of tumor suppressor genes and/or activation of oncogenes [9–11]. Specific CpG islands are commonly methylated in malignancies and the methylation pattern seems to be tumor type specific [3,9,11–15]. Epigenetic repression of the INK4a/ARF locus, encoding the tumor suppressors $p16^{INK4a}$ and $p14^{ARF}$, is a frequent event during immortalization of fibroblasts and epithelial cells [2,16,17]. In addition, hypomethylation of intragenic regions may result in derepression of transposable elements contributing to genomic instability [9].

Analysis of the impact of DNA methylation on processes relevant for cellular immortalization *in vitro* is complicated due to the fact that successive methylation changes may occur by time and number of population doublings (PDs). Long-term culture of fibroblasts and mesenchymal stromal cells is associated with specific senescence-associated DNA methylation changes [18]. In mesenchymal stromal cells, overexpression of TERT or immortalization with a doxycycline-inducible system (TERT and SV40-TAg) resulted in telomere extension but did not prevent senescence-associated DNA methylation [19]. It was also noted that methylation patterns were maintained throughout both long-term culture and aging but with highly significant differences at specific CpG sites [20]. However, for hematopoietic cells *in vitro* data are conflicting and limited to Epstein Barr Virus (EBV)-transformed lymphoblastoid B cell lines [21,22].

In the present study, genome-wide promoter-associated methylation was analyzed during spontaneous immortalization of T-cell cultures established from patients with Nijmegen breakage syndrome (NBS) and a healthy individual, using high-density arrays. A significant number of CpG site alterations throughout immortalization were shared with pediatric T-ALL suggesting a clinical relevance of these methylation changes.

Materials and Methods

T-cell Cultures and Culture Conditions

The studied T-cell cultures were established at the Sklodowska-Curie Memorial Cancer Center in Warsaw, Poland, and at Umeå University in Umeå, Sweden, using mitogen-initiated, Interleukin-2 (IL-2)–dependent cultures without genetic manipulations, as previously described [8,23,24]. The spontaneously immortalized T-cell lines (S3R, S4, and S9) were established from peripheral blood mononuclear cells derived from patients with NBS homozygous for the 657del5 mutation of the *NBS1* gene [8,23]. T-cell lines (L4 and L5) and their parental population (L2) as well as the primary T lymphoblast culture S1/PHA were derived from normal spleen [24]. The primary T-lymphoblast culture P7/R2 was derived from peripheral blood mononuclear cells of a healthy donor and was generated after initial 24-hour activation with 20 μg/ml Wheat Germ Agglutinin (WGA), followed by culture in standard medium without mitogen for the next 5 days and thereafter propagation in 20 U/ml of rIL-2 for 14 PDs. All cultures were maintained in standard medium [RPMI 1640, 10-12% fetal calf serum, 50 μg/ml gentamicin (Sigma-Aldrich, St Louis, MO)] supplemented with 20 U/ml rIL-2 (R&D Systems, Minneapolis, MN), in 5% CO_2 at 37°C. An approval from the Ethical Council in Warsaw, Poland, was obtained before collection of the NBS blood samples and the patients' guardians provided informed consent.

Cell cultures were grouped accordingly: primary with limited life-span *in vitro* (P7/R2 14 PDs, S1/PHA 2 PDs), pre-immortal (S3R 17 PDs, S4 12/18/48 PDs, S9 10 PDs, L2 5 PDs), and immortal (S3R 27/76/192 PDs, S4 68/223 PDs, S9 104 PDs, L4 195 PDs, L5 157 PDs). Pre-immortal and immortal T-cell cultures were separated by a period of growth crisis, in S3R at 21 to 25 PDs and in S4 at 62 to 67 PDs, but with no clear growth crisis period in cultures S9, L4, or L5.

T-ALL Samples

Diagnostic bone marrow samples from 43 pediatric T-ALL patients collected at the University Hospital in Umeå, Sweden, have been previously analyzed by the HumMeth27K (n = 43) and HumMeth450K (n = 10) Illumina methylation arrays (Illumina, San Diego, CA) and classified regarding CpG island methylator phenotype (CIMP) status [25]. Methylation array data from the HumMeth27K array was downloaded from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) database, GSE42079, and data from the HumMeth450K array has been deposited to the GEO database, GSE56070. The Regional Ethics Committee approved the study, and the patients and/or their guardians provided informed consent.

DNA Preparation

DNA was prepared with the Nucleon BACC2 kit (Amersham Biosciensces AB, Uppsala, Sweden) and DNA purity and concentration were determined by spectrophotometry (NanoDrop; Thermo Scientific, Wilmington, DE).

Genome-Wide CpG Site Methylation Profiling

The cell cultures were analyzed at different stages during immortalization using a high-density array covering 485,577 CpG sites (HumMeth450K, Illumina). The included CpG sites are located in different genomic regions, but the main focus is promoter-associated regions and CpG islands. The array definition of genomic regions and the relations to CpG islands are shown in Supplementary Figure S1. A methylated control sample (Human HCT116 DKO methylated DNA; Zymo Research, Irvine, CA, USA) enzymatically methylated on all cytosines by M.SssI methyltransferase and a non-methylated (<5%) control double knocked out for DNA methyltransferases (DNMT1 –/– and DNMT3B –/–; Human HCT116 DKO non-methylated DNA; Zymo Research) were included. For each sample, 500 ng of DNA was bisulfite converted with the EZ-96 DNA Methylation-Gold Kit (Zymo Research) according to the manufacturer's manual. Two hundred nanograms of bisulfite-converted DNA was

applied to each array, which was handled according to the Illumina provided protocol and scanned with an iScan SQ instrument (Illumina). The fluorescence intensities were extracted using the Methylation module (1.9.0) in the Genome Studio software (V2011.1).

Pre-Processing of Methylation Array Data

CpGs on the X and Y chromosomes were omitted from the analysis to avoid gender-related methylation biases. In addition, CpG sites lacking observations due to less than or equal three reported beads/array or CpG sites with low detection *P* value (>.05) were excluded. Likewise, CpG sites located at or close to a (10-bp) single nucleotide polymorphism (SNP) as well as CpG sites located in intergenic regions were excluded. All filtration steps are shown in Supplementary Figure S2. The methylation levels for the remaining 330,354 CpG sites were determined by calculating the ratio (i.e., the β value) between the fluorescent intensity from the methylated alleles and the total intensity, as defined in the Genome Studio software (Illumina). The β value ranges in theory from 0, corresponding to completely unmethylated DNA, to 1, representing fully methylated DNA. To compensate for the two different bead types used in the HumMeth450K array, the β values were normalized using the BMIQ method [26,27].

The methylation array data have been deposited to the NCBI GEO database, GSE56070.

Verification of Methylation Array Data by Pyrosequencing

DNA samples were sent to the Genome Centre Queen Mary, University of London for targeted pyrosequencing on a selection of genomic regions overlapping with specific CpGs in the methylation array, including TAL1 (cg19797376), KLF4 (cg07309102), HOXD8 (cg15520279), and TWIST1 (cg24446548). Pyrosequencing was performed according to the manufacturer's protocol by bisulfite treatment of DNA (EZ DNA Methylation; Zymo Research), followed by polymerase chain reaction (PCR) amplification and pyrosequencing using PyroMark Gold Q96 Reagents (Qiagen, Sollentuna, Sweden) in the PSQ 96MA instrument with PSQ 96MA software V2.1 (Qiagen). The following set of primers were used: TAL1_F: ATGGGGGTTAGAGAGAGAATGA; TAL1_R:ACCTCCTCAACCAAATCTC; TAL1_seq: GGGGGATTT TAAGGT; HOXD8_F: AGTGATAGTAGTAGTAAGTGGGATT GAT; HOXD8_R: AACAACCCCCCCACAAACCCC; HOXD8_seq: GTTTTGTATTTGGAGTATAG; KLF4_F: AGGTTGTAGAGAAG GAAGTTATAAGTAAG; KLF4_R: CAACAACCTCCCCCACCAC TAT; KLF4_seq: ATACCCCCAAATAAAACTAACTAC; TWIST1_F: GGAGGTATAAGAGTTTTTAAGTTTGTAG; TWIST1_R: ACACCCCCCCAAACCTCCTA; TWIST1_seq: AGAGTTTT TAAGTTTGTAGTT.

RNA Preparation and Pre-Processing of Gene Expression Data

Total RNA was isolated using TRIzol reagent (Invitrogen, Stockholm, Sweden) according to the manufacturer's protocol. The RNA quality was analyzed in a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA) and RNA integrity number was >9 in all samples. In brief, 200 ng of total RNA was used for cRNA production by the Illumina TotalPrep RNA amplification kit (Ambion Inc, St Austin, TX) according to the provided protocol. The biotin-labeled cRNA was purified and the quality was evaluated using the RNA 6000 pico kit in the 2100 Bioanalyzer (Agilent Technologies).

A total of 750 ng of biotinylated cRNA was hybridized to the human HT12 Illumina Beadchip gene expression array (Illumina) according to the manufacturer's protocol and scanned using the Illumina Bead Array Reader (Illumina). Illumina Genome Studio software (V2011.1) with gene expression module (1.9.0) was used for data extraction and normalization using the rank invariant normalization. Selected parts of the gene expression array data have been previously published [8]. Gene expression array data have been deposited to the NCBI GEO database, GSE56070.

Quantitative Reverse Transcription–PCR Analysis

cDNA was prepared by reverse transcription (RT) of 500 ng of total RNA with the Superscript II Reverse Transcriptase kit (Invitrogen) together with random hexamers (Applied Biosystems, Inc, Foster City, CA) and RNasin (Promega, Nacka, Sweden) according to the manufacturer's instructions (Invitrogen).

Expression levels of selected genes were determined by quantitative PCR in duplicates and a standard curve was included in each assay to monitor PCR efficiency. The following genes were analyzed by TaqMan assays on demand according to the manufacturer's protocol using the TaqMan Universal PCR Mastermix in the ABI PRISM 7900HT Instrument (Applied Biosystems, Inc): *TATA-binding protein gene* (*TBP*) (Hs99999910_m1), *BMI1 polycomb ringfinger oncogene (BMI1)* (Hs00180411_m1), *Chromobox homolog 2 (CBX2)* (Hs01034268_m1), *chromobox homolog 7 (CBX7)* (Hs00545603_s1), *Enhancer of Zeste homolog 2 (EZH2)* (Hs01016789_m1), and *supressor of Zeste 12 (SUZ12)* (Hs00248742_m1). Relative mRNA levels were normalized to a housekeeping *TBP gene* and fold change was calculated by the $2^{-\Delta\Delta Ct}$ method using P7/R2 and S1/PHA as primary mortal T-cell culture references [28].

Bioinformatic and Statistical Analyses

Downstream analyses of the methylation and gene expression array data were performed using R (v2.15.0). The CpG sites were matched to genes on the expression array according to their RefSeq accessions. Gene annotations from both gene expression and methylation arrays were matched to the hg19 assembly (NCBI) and their annotations were updated. Genes with discontinued accessions or non-consistent annotations were excluded from further analysis. In the downstream analysis, the normalized β values and the $\log_2$-transformed normalized signal intensities were used as measures of methylation and gene expression levels, respectively.

The primary T-lymphoblast culture P7/R2 was used as a universal reference and the methylation pattern for each time point of the two cell cultures was analyzed in relation to P7/R2 (Figure 1). The CpG sites were classified as *de novo* altered CpGs if the difference in β values was greater than 0.4 (gain of methylation) or less than −0.4 (loss of methylation). Although the methylation of the P7/R2 cell line was highly correlated with the pre-crisis samples in each cell line, it was excluded from further analysis of differential methylation and expression in the S3R and S4 cell cultures.

For each CpG site, the alteration between post-crisis and pre-crisis samples was measured with the maximum differences in methylation ($\delta\beta$). CpG sites with an $\delta\beta$ greater than 0.4 or less than −0.4 were classified as differentially methylated (DM-CpG). For gene expression, genes expressed at background level were censored to the largest of the 95th percentiles of the negative controls across the arrays. The alteration between post-crisis and pre-crisis samples was then measured with the maximum differences in $\log_2$-transformed censored intensities (*M*). Genes with an *M* value greater than $\log_2$ (1.7) or less than $-\log_2(1.7)$, corresponding to a fold change of 1.7 and expressed above background level were classified as differentially

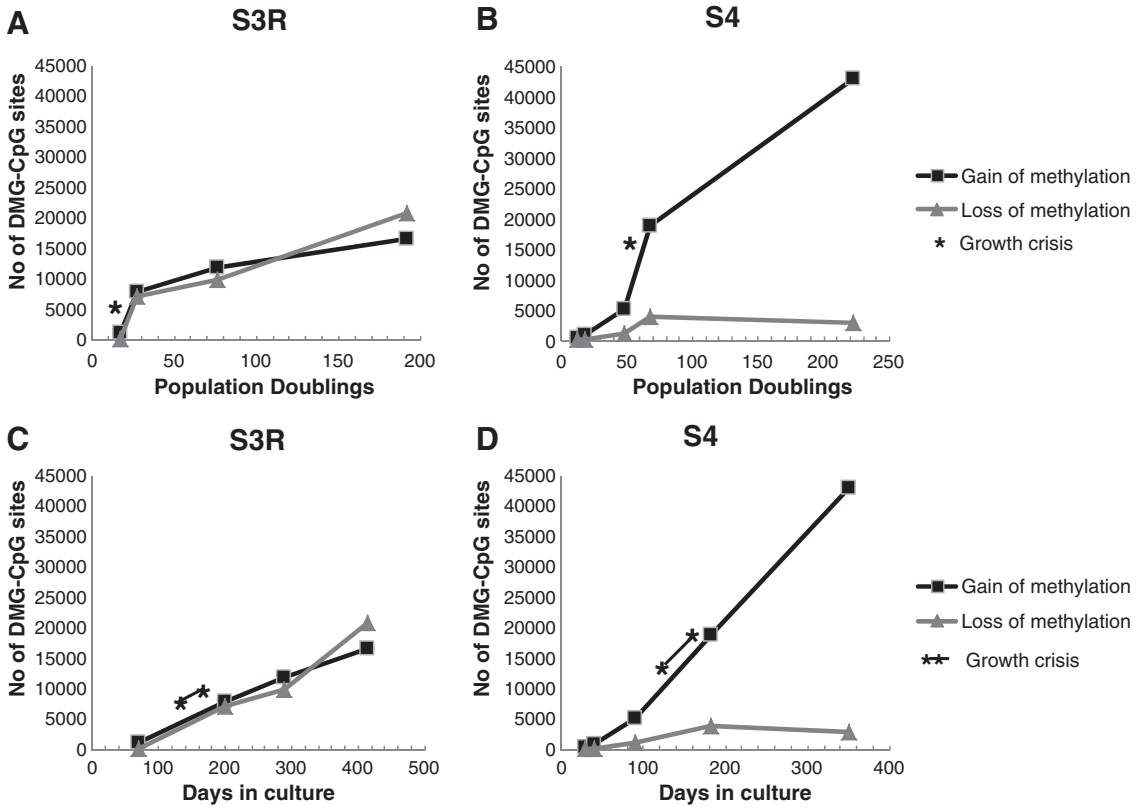


Figure 1. Methylation alterations accumulated during the immortalization process. The number of *de novo* altered CpG sites at increasing PDs (A and B) or days in culture (C and D) in S3R (A and C) and S4 (B and D) is shown. The gain of methylation is defined as $\delta\beta > 0.4$, and loss of methylation is defined as $\delta\beta < -0.4$, compared with a stimulated primary T-cell culture (P7/R2). The growth crisis periods are marked with asterisks.

expressed genes (DEG). Further details of the analysis can be found in Supplementary Figure S2. Differently methylated gene (DMG) and DEG found in both cell cultures (DMG/DEG) were selected for further analysis (Figure 2). The significance of the overlaps between the two cell cultures was determined using the chi-square test.

The significance of the overlap between DMG and DEG (Figure 2) and distributions of downregulated genes among methylated and demethylated genes (Table 2) were evaluated using a permutation-based test. While keeping the gene-CpG structure intact, the gene expression profiles were randomly assigned to genes 2000 times. Thus, the distributions under the null hypothesis of independence between methylation and gene expression and *P* values for the size of the overlap and the distribution of downregulated genes were obtained.

To analyze common methylation patterns between the two T-cell cultures and T-ALL samples, hierarchical clustering with Euclidean distance matrix using the Ward method [29] was performed and visualized in a heat map. Publicly available sorted T-cells (CD3+) and hematopoietic stem cells (CD34+CD38−) from healthy donors were used as controls. The methylation data from the sorted cells were previously published by the Cancer Genome Atlas Research Network [30] and were downloaded from the NCBI GEO database, GSE49618.

The polycomb target gene (PCTG) lists from Lee et al. [31] were compared with the genes with common methylation changes between T-ALL samples and T-cell cultures. We hypothesized that the proportion of PCTGs in the list of shared changed CpG sites was larger than expected by chance and we tested our hypothesis using a binominal test. The systems biology tool Metacore from GeneGo Inc (St Joseph, MI) was used to identify networks and processes of possible relevance for immortalization.

Results

Quality Control and Reproducibility of Methylation Array Data

The quality of each individual array was evaluated with the built-in controls, i.e., bisulfite conversion, staining, negative controls, hybridization, and specificity. A replicate sample was included in the HumMeth450K array to assess inter-assay reproducibility (R^2 = 0.989). Selected genomic regions for TAL1 (cg19797376), KLF4 (cg07309102), TWIST1 (cg24446548), and HOXD8 (cg15520279) were separately analyzed by pyrosequencing at increasing PDs in the S3R and S4 cell cultures and compared with CpG site data from the HumMeth450K array, showing a strong correlation between the methods (Supplementary Figure S3).

DNA Methylation Alterations during T-Cell Immortalization

Genome-wide promoter methylation status was determined by the HumMeth450K array at different stages during spontaneous immortalization of the two human T-cell cultures, S3R and S4. Both cultures experienced a growth crisis period, between 21 and 25 PDs (93 days) in S3R and between 62 and 67 PDs (45 days) in S4. After pre-processing of the array data, as detailed in the Materials and Methods section and in Supplementary Figure S2, 330,354 CpG sites

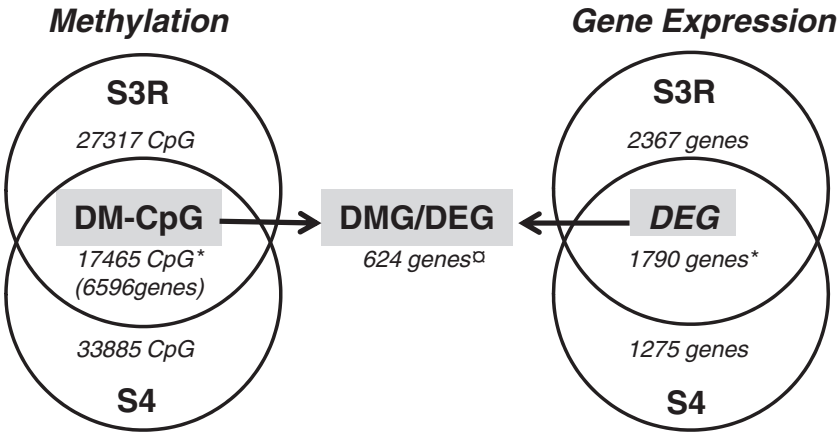


Figure 2. Commonly DM-CpG sites and differentially expressed genes in immortal T-cell cultures. (A) The DM-CpG sites ($\delta\beta > 0.4$ or <-0.4) and the DMG (fold change ± 1.7) pre-crisis *versus* post-crisis in the S3R and S4 cell cultures were combined in Venn diagrams. The commonly DM-CpG sites (17,465 CpGs representing 6596 genes) and DEG (1790 genes) are highlighted in the figure as well as the combined DMG/DEG list (624 genes). There was a significantly higher proportion of common DM-CpG and DEG than expected by random (*$P < .001$) in the immortal cultures. In contrast, the number of genes that were both DMG and DEG was significantly smaller than expected (¤$P < .001$).

remained for analysis. The methylation status of these CpGs were analyzed at increasing PDs and compared with a primary stimulated T-cell culture (P7/R2). Both the S3R and S4 cultures showed an accumulation of *de novo* altered CpG sites with increasing PDs (Figure 1, *A* and *B*). The most pronounced change was observed in cultures escaping from the growth crisis period (S3R 27 PDs and S4 68 PDs; Figure 1, *A* and *B*). However, the growth crisis period lasted for a long period of time and when analyzing methylation changes in relation to days in culture the rate of changes appeared rather constant (Figure 1, *C* and *D*). The fraction of *de novo* methylated sites was, in S3R, 0.3% at 17 PDs (pre-crisis) and 5% at 192 PDs (post-crisis) and, in S4, 0.1% at 12 PDs (pre-crisis) and 13% at 223 PDs (post-crisis) of all CpG sites analyzed and in relation to the primary stimulated T-cell culture (P7/R2). Loss of methylation was less common in S4 in comparison to the S3R culture, in which gain and loss of methylation were equally common (Figure 1, *A–D*).

Next, by filtering for shared DM-CpG sites (demethylated or methylated) between the cultures pre-crisis *versus* post-crisis, a significantly larger overlap ($P < .001$) than expected by random was observed. This overlap consisted of 17,465 DM-CpGs with potential importance for immortalization (Figure 2). By focusing on these DM-CpGs, it was further shown that the unmethylated CpG sites close to transcription start sites (TSSs) (TSS200/exon 1) and CpG islands (islands/shores) were preferentially and gradually methylated during the process (Table 1 and Supplementary Figure S4). Demethylation was observed in methylated regions far away from CpG islands, here titled "open sea" and "shelf" (Table 1 and Supplementary Figure S4).

Table 1. Preferential Methylation in CpG Islands and Regions around TSS

	S3R				S4				
(PD)	17	27	76	192	12	18	48	68	223
Island	0.18	0.41	0.57	0.70	0.18	0.20	0.30	0.54	0.84
Shore	0.33	0.45	0.53	0.62	0.32	0.34	0.39	0.52	0.73
Shelf	0.58	0.40	0.37	0.37	0.57	0.56	0.48	0.41	0.47
Open sea	0.57	0.38	0.34	0.35	0.57	0.55	0.48	0.40	0.47

	S3R				S4				
(PD)	17	27	76	192	12	18	48	68	223
TSS1500	0.32	0.43	0.51	0.60	0.31	0.32	0.37	0.51	0.71
TSS200	0.24	0.39	0.53	0.63	0.23	0.25	0.33	0.52	0.79
5'UTR	0.30	0.39	0.48	0.59	0.29	0.31	0.35	0.48	0.72
1st exon	0.26	0.40	0.52	0.60	0.27	0.28	0.33	0.50	0.76
Gene body	0.42	0.41	0.46	0.51	0.42	0.42	0.42	0.47	0.63
3'UTR	0.48	0.42	0.44	0.45	0.48	0.48	0.45	0.47	0.58

1.00 0.90 0.80 0.70 0.60 0.50 0.40 0.30 0.20 0.10 0.00

Average methylation (average β) of the commonly differently methylated 17,465 CpG sites for the respective genomic region and CpG island location is shown at increasing PDs. The colors represent a scale from unmethylated (green) to fully methylated (red) as shown in the figure.

Table 2. Decreased Gene Expression Was Associated with Increased Methylation around TSS and Decreased Methylation in Gene Body

Genomic Region	Methylated	Demethylated
TSS1500	61*	52
TSS200	70***	50
5′UTR	68***	53
First exon	60	0***[1]
Body	58	69***
3′UTR	64	79***

The percentage of downregulated genes in gene regions that were methylated/demethylated is shown. For the significance based on permutation of gene expression probes, expected percentage of downregulated genes is 50% to 51% in all categories. The level of significance is represented by $^{*}P < .05$, $^{**}P < .01$, and $^{***}P < .001$; 1 denotes categories with less than five observations.

Genomic Distribution of Coinciding DM-CpG Sites and Differently Expressed Genes

In addition to the DM-CpGs, a significant overlap ($P < .001$) of 1790 differently expressed genes (DEG) either upregulated or downregulated pre-crisis *versus* post-crisis was identified, indicating shared pathways for immortalization (Figure 2). A gene with at least one DM-CpG was defined as a DMG. The DMG and the DEG lists were combined and an overlap of 624 genes both differentially methylated and differentially expressed was identified (DMG/DEG). The DMG/DEG list was analyzed for associations to methylation in different genomic regions. A significant overrepresentation of downregulated genes was observed when methylation occurred close to TSS ($P < .001$; Table 2). In contrast, decreased methylation within the gene body region was associated with decreased expression ($P < .001$; Table 2). However, the DMG/DEG overlap was significantly lower than expected by random ($P < .001$), indicating that the majority of methylation alterations have little or no effect on gene expression levels (Figure 2).

Functional Analysis of Shared DM-CpGs in Immortal T-Cell Cultures and T-Cell Leukemia

To identify methylation alterations of potential importance for both *in vitro* T-cell immortalization and *in vivo* malignancy, we applied data from 10 diagnostic pediatric T-ALL samples (7 CIMP+ and 3 CIMP−) on a heat map showing the 17,465 commonly altered CpGs in the immortal S3R and S4 T-cell cultures (Figures 2 and 3*A*). Sorted cells from healthy donors were used as controls [30]. CD34+/CD38− hematopoietic stem cells represented immature T-cells and CD3+ represented mature T-cells. Immature and mature T-cells showed similar methylation profiles as pre-immortal cell cultures regarding the CpG sites that were commonly altered during the immortalization process (Figure 3*A*).

Hierarchical clustering of the cell cultures and leukemias separated the CpGs into three clusters; cluster 1 in which *de novo* methylated CpGs in the immortal cells overlapped with methylated CpGs in CIMP+ T-ALL diagnostic samples. CpGs in cluster 2 were *de novo* methylated in immortal cell cultures but less methylated in leukemia, and cluster 3 demethylated in immortal cell cultures but methylated in leukemia (Figure 3*A*). Bioinformatic analysis revealed a significant overrepresentation ($P < .001$) of PCTGs [31] compared to random methylation (Figure 3*A*) in all clusters, but the overrepresentation was most evident in cluster 1 where 51% of the CpG sites were located in PCTGs compared to 23% and 21%, respectively, in clusters 2 and 3. Furthermore, the majority (62%) of the shared CpG sites in immortal cell cultures and CIMP+ leukemias were located in CpG islands.

To verify the overlap in altered genes in immortal T-cell cultures and CIMP+ T-ALL, data from 43 diagnostic T-ALL samples analyzed on the HumMeth27K array as well as three additional cell cultures derived from one patient with NBS and one healthy individual were included; 1478 CpG sites of the 17,465 shared DM-CpGs (Figure 2) were present in the HumMeth27K array. Cell cultures and leukemias were clustered, and CpGs sorted in the three clusters were identified in Figure 3*A*. All immortal cell cultures showed similar alterations and cluster 1 was commonly methylated in immortal cell cultures and CIMP+ leukemias (Figure 3*B*).

To further investigate the cellular pathways commonly altered in immortal cell cultures and leukemia, the list of shared DM-CpGs in cluster 1 (Figure 3*A*) was applied to Metacore process network analysis (Figure 4, *A*–*C*). The resulting *P* values were compared at increasing PDs in cell cultures, in grouped CIMP+ ($n = 7$) and CIMP− ($n = 3$) leukemias, and the list of PCTGs [31] (Figure 4, *A* and *B*). This comparison allowed assessment of the number of alterations for each individual process network with corrections for pathway and gene list sizes. The process networks of highest relevance were cell adhesion, cytoskeletal remodeling, development, and signal transduction processes (Figure 4*B*). Interestingly, a large set of genes in the Wnt signaling pathway became methylated during immortalization of cell cultures and in diagnostic CIMP-positive leukemias (Figure 4*C*).

To identify potentially deregulated components/pathways in the methylation and chromatin modifying machinery during immortalization, we analyzed gene expression of DNA methyltransferases (*DNMT1*, *DNMT3A*, and *DNMT3B*) and *polycomb repressive complex 1* (*PRC1*) and *PRC2* genes in the S3R and S4 cell cultures. PRC1/2 was significantly changed regarding the *EZH2* (S3R, S4), *BMI1* (S4), *PCGF2* (S3R, S4), *CBX2* (S3R, S4), and *CBX6* (S3R, S4) genes, all showing up-regulation. In contrast, decreased levels of *CBX4* (S3R) and *CBX7* (S3R, S4) were recorded. In S3R, *DNMT1* remained stable throughout the culture, whereas *DNMT3A* and *DNMT3B* decreased (Supplementary Figure S5). In contrast, S4 showed increased expression of *DNMT1* and stable expression of *DNMT3A* and *3B* (Supplementary Figure S5).

Discussion

Here, we have shown that spontaneous immortalization of T-cell cultures was associated with common DNA methylation alterations found to be shared at a high frequency with CIMP+ T-ALL diagnostic samples, indicating that the *in vitro* established methylation alterations might be relevant in the clinical setting.

In our study, two IL-2–dependent T-cell cultures were analyzed with a high-density genome-wide methylation array at several time points, from primary culture, over a growth crisis period, and until clonal immortal cell cultures emerged (approximately 350-400 days). We have previously shown that the initial polyclonal cultures gradually became monoclonal indicating a continuous loss of cells likely due to the senescence program [8,23]. Methylation changes were recorded in pre-immortal S4 cells already before the growth crisis period, which might have contributed to senescence bypass/growth crisis escape. The accumulation of methylation alterations was rather constant when analyzing changes against days in culture. In contrast, when analyzed in relation to PDs, the changes appeared more stepwise. Stepwise methylation changes have been recorded during immortalization of human mammary epithelial cell cultures in which a "stasis" (stress-induced senescence barrier) and a "telomere dysfunction" (i.e., crisis) barrier were identified [3,32]. The

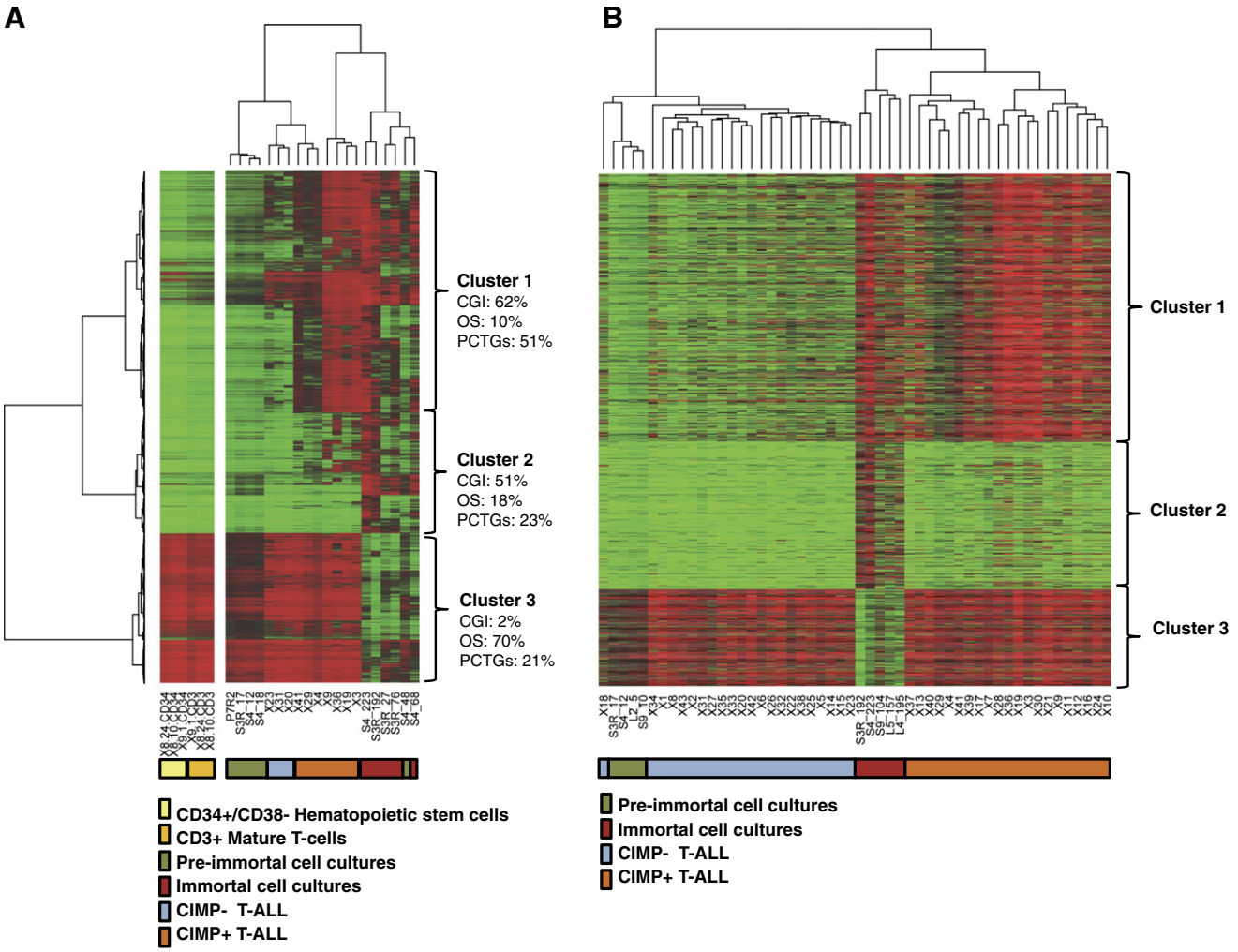


Figure 3. Shared methylation alterations in immortal T-cell cultures and diagnostic CIMP + T-ALL samples. (A) Methylation HumMeth450K array data for the commonly altered 17,465 CpGs in T-cell cultures visualized in a heat map together with 10 diagnostic T-ALL samples. Separated CD34 +/CD38 − (hematopoietic stem cells, $n = 3$) and CD3 + (mature T-cells, $n = 3$) cells from healthy donors were included as controls. The CpG sites were separated into three distinct clusters. The percentage of CpG sites located within CpG islands (CGI) and open sea (OS) regions in each cluster is displayed as well as the percentage of PCTGs in each cluster defined by Lee et al. [31]. The diagnostic T-ALL sample CIMP status is shown in the figure as well as the pre-immortal or post-immortal status of the cell cultures. (B) Verification of shared altered CpG sites in immortal cell cultures and in CIMP + T-ALL leukemias by HumMeth27K array analysis; 1478 CpGs of the 17,465 CpGs commonly altered CpG sites in immortal S3R and S4 cultures were present in the 27K array and used to verify data in a larger number of T-ALL samples and in three additional cell cultures derived from one patient with NBS (S9) and one healthy individual (L4 and L5).

immortalized human mammary epithelial cells were experimentally achieved and thereby not comparable to the spontaneously immortalized T-cells analyzed in the present study.

Both T-cell cultures studied showed an accumulation of *de novo* altered CpG sites during immortalization, with similar levels of gain and loss of methylation for S3R while culture S4 was dominated by gain of methylated CpG sites. Interestingly, the two cultures shared a large fraction of DM-CpG sites between pre-crisis and post-crisis cells. For these sites, a preferential gain of methylation in CpG islands and CpG sites located close to TSS was observed, whereas loss of methylation preferentially occurred in gene body regions and at sites located far away from CpG islands ("open sea"). Genomic regions close to TSS are known to contain a high density of CpG islands, whereas the gene body has a lower density, explaining the overlapping results of genomic regions and CpG island relations [9]. The chromosomal distribution of the methylation alterations paralleled the distribution of CpG sites on the array, and methylation alterations therefore appeared as a genome-wide phenomenon (data not shown).

The collected findings of common and non-random methylation changes in the immortalized T-cells indicate a functional meaning. One question was to what extent the methylation pattern was coupled to gene expression. To answer this question, we combined methylation data with our previously published gene expression array data [8]. Methylation and gene expression did not correlate for the majority of genes, indicating that many methylation alterations did not relate to gene expression. However, we could identify genes with strong negative or positive correlation to methylation status. The direction of the correlation seemed to depend on where in the promoter region the methylation occurred. Generally, silenced gene expression was significantly associated with increased methylation close to TSS and loss of methylation in the gene body region. However, increased methylation within the gene body was observed as well and might reflect

Figure 4. Ontology analysis of common cellular processes affected by methylation. Metacore process network analysis of the shared 8255 CpG sites (cluster 1; Figure 3*A*) between post-crisis T-cell cultures and diagnostic CIMP + T-ALL (A). CIMP − ($n = 3$) and CIMP + ($n = 7$) T-ALL samples were grouped in the analysis, whereas each time point is shown for the cell cultures. A significant process ($P < .05$) is indicated in red in the heat map. (B and C) The top processes in the Metacore process network analysis are shown as well as the WNT signaling genes representing a process with many affected genes. The red thermometers represents (1) polycomb target genes and (2) shared 8255 CpG sites in post-crisis T-cell cultures/CIMP + T-ALL.

the presence of enhancers or regulatory regions as recently suggested by Varley et al. [33].

To identify methylation alterations of potential importance for both *in vitro* T-cell immortalization and *in vivo* malignancy, we applied data from diagnostic pediatric T-ALL samples to the analysis. T-ALL originates from precursor cells in the bone marrow, whereas our T-cell culture models are derived from mature peripheral blood lymphocytes. However, the methylation profiles of the CpG sites that

were differently methylated during the immortalization process were comparable between immature and mature cells from healthy donors and pre-immortal cell cultures and confirm our T-cell culture model as suitable for comparable analysis with T-ALL.

We have recently identified methylation alterations within T-ALL samples separating them into subgroups (CIMP+/CIMP−) with different prognosis [25]. Interestingly, a substantial fraction of the altered CpG sites observed in the T-cell cultures had similar methylation profiles in the CIMP+ T-ALL samples. This commonality indicates that the *in vitro* established methylation changes were relevant also in the clinical setting. The affected CpG sites were not unbiased regarding the function of associated genes since differential methylation of PCTGs was overrepresented. Furthermore, bioinformatic analyses showed a pattern of overlapping networks among the affected genes in post-crisis T-cell cultures, CIMP+ leukemias, and the PCTG list [31]. These shared processes were involved in cell adhesion, cytoskeleton remodeling, development, and signal transduction. Of special notice was a high number of methylation altered genes in the Wnt signaling pathways. The reason for this is unclear but indicates selectivity in the methylation targets.

Gene expression alterations of DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B) as well as subunits in the PRC1/2 were observed in both cultures suggesting a functional role for the observed methylation changes during the immortalization process. Previous observations have shown that different DNA methylation changes typical for cancer cells may evolve from senescence-related alterations triggered by different stimuli [34] and altered DNA methylation acquired in senescent cells can be retained when these cells bypass senescence [35].

The three spontaneously immortalized cell cultures, S3R, S4, and S9, were derived from individuals diagnosed with NBS with increased risk of leukemia due to DNA repair deficiency. However, the patients were not diagnosed with any malignancy at the time when the cultures were established. The observed methylation alterations in the immortal NBS-derived T-cell cultures overlapped with T-cell cultures derived from a healthy individual (L4 and L5) indicating general accumulated methylation alterations during the immortalization process.

In conclusion, our analysis identified a high number of commonly methylated CpG sites in immortalized T-cell cultures and diagnostic CIMP+ T-ALL samples with potential significance for malignant transformation. Whether CIMP+ and CIMP− leukemias represent subtypes with different routes for transformation has to be further investigated. One alterative hypothesis is that CIMP+ and CIMP− leukemias reflect cells with diverse replicative histories, where CIMP+ leukemias might have undergone a large number of replication rounds and thereby demonstrate DNA methylation pattern similar to long-term cultured T-cells. These issues have to be further evaluated to better understand the relevance of methylation during immortalization and the development of hematological malignancies.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neo.2014.07.001.

Acknowledgments

This work was supported by grants from the Swedish Cancer Society (G.R.), the Swedish Research Council (G.R.), Dnr 340-2013-5185 (P.R.), the Swedish Childhood Cancer Foundation (G.R.), the Medical Faculty, Umeå University (G.R.), and Lion's Cancer Research Foundation, Umeå University (S.D. and G.R.). The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2011 under grant agreement No. 200950 (G.R.).

References

[1] Fridman AL and Tainsky MA (2008). Critical pathways in cellular senescence and immortalization revealed by gene expression profiling. *Oncogene* **27**, 5975–5987.

[2] Benanti JA, Wang ML, Myers HE, Robinson KL, Grandori C, and Galloway DA (2007). Epigenetic down-regulation of ARF expression is a selection step in immortalization of human fibroblasts by c-Myc. *Mol Cancer Res* **5**, 1181–1189.

[3] Novak P, Jensen TJ, Garbe JC, Stampfer MR, and Futscher BW (2009). Stepwise DNA methylation changes are linked to escape from defined proliferation barriers and mammary epithelial cell immortalization. *Cancer Res* **69**, 5251–5258.

[4] Takai D and Jones PA (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A* **99**, 3740–3745.

[5] Liu C, Fang X, Ge Z, Jalink M, Kyo S, Bjorkholm M, Gruber A, Sjöberg J, and Xu D (2007). The *telomerase reverse transcriptase* (*hTERT*) gene is a direct target of the histone methyltransferase SMYD3. *Cancer Res* **67**, 2626–2631.

[6] Renaud S, Loukinov D, Abdullaev Z, Guilleret I, Bosman FT, Lobanenkov V, and Benhattar J (2007). Dual role of DNA methylation inside and outside of CTCF-binding regions in the transcriptional regulation of the telomerase hTERT gene. *Nucleic Acids Res* **35**, 1245–1256.

[7] Zinn RL, Pruitt K, Eguchi S, Baylin SB, and Herman JG (2007). hTERT is expressed in cancer cell lines despite promoter DNA methylation by preservation of unmethylated DNA and active chromatin around the transcription start site. *Cancer Res* **67**, 194–201.

[8] Degerman S, Siwicki JK, Osterman P, Lafferty-Whyte K, Nicol Keith W, and Roos G (2010). Telomerase upregulation is a postcrisis event during senescence bypass and immortalization of two Nijmegen breakage syndrome T cell cultures. *Aging Cell* **9**, 220–235.

[9] McCabe MT, Brandes JC, and Vertino PM (2009). Cancer DNA methylation: molecular mechanisms and clinical implications. *Clin Cancer Res* **15**, 3927–3937.

[10] Boultwood J and Wainscoat JS (2007). Gene silencing by DNA methylation in haematological malignancies. *Br J Haematol* **138**, 3–11.

[11] Esteller M (2008). Epigenetics in cancer. *N Engl J Med* **358**, 1148–1159.

[12] Costello JF, Frühwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomaki P, and Lang JC, et al (2000). Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* **24**, 132–138.

[13] Ehrich M, Turner J, Gibbs P, Lipton L, Giovanneti M, Cantor C, and van den Boom D (2008). Cytosine methylation profiling of cancer cell lines. *Proc Natl Acad Sci U S A* **105**, 4844–4849.

[14] Martin-Subero JI, Ammerpohl O, Bibikova M, Wickham-Garcia E, Agirre X, Alvarez S, Bruggemann M, Bug S, Calasanz MJ, and Deckert M, et al (2009). A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms. *PLoS One* **4**, e6986.

[15] Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Gayther SA, Apostolidou S, Jones A, Lechner M, Beck S, and Jacobs IJ, et al (2009). An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One* **4**, e8274.

[16] Shao G, Balajee AS, Hei TK, and Zhao Y (2008). p16INK4a downregulation is involved in immortalization of primary human prostate epithelial cells induced by telomerase. *Mol Carcinog* **47**, 775–783.

[17] Tzatsos A, Pfau R, Kampranis SC, and Tsichlis PN (2009). Ndy1/KDM2B immortalizes mouse embryonic fibroblasts by repressing the Ink4a/Arf locus. *Proc Natl Acad Sci U S A* **106**, 2641–2646.

[18] Koch CM, Joussen S, Schellenberg A, Lin Q, Zenke M, and Wagner W (2012). Monitoring of cellular senescence by DNA-methylation at specific CpG sites. *Aging Cell* **11**, 366–369.

[19] Koch CM, Reck K, Shao K, Lin Q, Joussen S, Ziegler P, Walenda G, Drescher W, Opalka B, and May T, et al (2013). Pluripotent stem cells escape from senescence-associated DNA methylation changes. *Genome Res* **23**, 248–259.

[20] Bork S, Pfister S, Witt H, Horn P, Korn B, Ho AD, and Wagner W (2010). DNA methylation pattern changes upon long-term culture and aging of human mesenchymal stromal cells. *Aging Cell* **9**, 54–63.

[21] Grafodatskaya D, Choufani S, Ferreira JC, Butcher DT, Lou Y, Zhao C, Scherer SW, and Weksberg R (2010). EBV transformation and cell culturing destabilizes DNA methylation in human lymphoblastoid cell lines. *Genomics* **95**, 73–83.

[22] Zhang Z, Liu J, Kaur M, and Krantz ID (2012). Characterization of DNA methylation and its association with other biological systems in lymphoblastoid cell lines. *Genomics* **99**, 209–219.

[23] Siwicki JK, Degerman S, Chrzanowska KH, and Roos G (2003). Telomere maintenance and cell cycle regulation in spontaneously immortalized T-cell lines from Nijmegen breakage syndrome patients. *Exp Cell Res* **287**, 178–189.

[24] Siwicki JK, Hedberg Y, Nowak R, Lodén M, Zhao J, Landberg G, and Roos G (2000). Long-term cultured IL-2-dependent T cell lines demonstrate $p16^{INK4a}$ overexpression, normal pRb/p53, and upregulation of cyclins E or D2. *Exp Gerontol* **35**, 375–388.

[25] Borssén M, Palmqvist L, Karrman K, Abrahamsson J, Behrendtz M, Heldrup J, Forestier E, Roos G, and Degerman S (2013). Promoter DNA methylation pattern identifies prognostic subgroups in childhood T-cell acute lymphoblastic leukemia. *PLoS One* **8**, e65373.

[26] Marabita F, Almgren M, Lindholm ME, Ruhrmann S, Fagerström-Billai F, Jagodic M, Sundberg CJ, Ekstrom TJ, Teschendorff AE, and Tegnér J, et al (2013). An evaluation of analysis pipelines for DNA methylation profiling using the Illumina HumanMethylation450 BeadChip platform. *Epigenetics* **8**, 333–346.

[27] Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D, and Beck S (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196.

[28] Schmittgen TD and Livak KJ (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101–1108.

[29] Ward Jr JH (1963). Hierarchical grouping to optimize an objective function. *J Am Stat Assoc* **58**, 236–244.

[30] Cancer Genome Atlas Research Network (2013). Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059–2074.

[31] Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, and Isono K, et al (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313.

[32] Garbe JC, Bhattacharya S, Merchant B, Bassett E, Swisshelm K, Feiler HS, Wyrobek AJ, and Stampfer MR (2009). Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res* **69**, 7557–7568.

[33] Varley KE, Gertz J, Bowling KM, Parker SL, Reddy TE, Pauli-Behn F, Cross MK, Williams BA, Stamatoyannopoulos JA, and Crawford GE, et al (2013). Dynamic DNA methylation across diverse human cell lines and tissues. *Genome Res* **23**, 555–567.

[34] Decottignies A and d'Adda di Fagagna F (2011). Epigenetic alterations associated with cellular senescence: a barrier against tumorigenesis or a red carpet for cancer? *Semin Cancer Biol* **21**, 360–366.

[35] Cruickshanks HA, McBryan T, Nelson DM, Vanderkraats ND, Shah PP, van Tuyn J, Singh Rai T, Brock C, Donahue G, and Dunican DS, et al (2013). Senescent cells harbour features of the cancer epigenome. *Nat Cell Biol* **15**, 1495–1506.