UNIVERSITY OF
LEICESTER

Epithelial-to-mesenchymal transition transcription factors (EMT-TFs) induce genomic instability via down-regulation of DNA ligase I (LIG1) in carcinoma cell lines

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

by

Noura Naji Alibrahim, BSc MSc

Leicester Cancer Research Centre
Department of Genetics and Genome Biology
University of Leicester

2019

Epithelial-to-mesenchymal transition transcription factors (EMT-TFs) induce genomic instability via down-regulation of DNA ligase I (LIG1) in carcinoma cell lines

Noura Naji Alibrahim

Abstract:

Epithelial-to-mesenchymal transition (EMT) is the reversible process in which cells lose their polarity and cell-cell adhesion and acquire mesenchymal traits. EMT is induced by a network of transcription factors known as EMT-TFs. This group of proteins includes members of ZEB, TWIST and SNAIL families; they play a role in embryonic development, and also in pathological conditions, tissue fibrosis and cancer. EMT-TFs control various cancer related pathways including cell cycle control, DNA damage response and chemo- and radioresistance. Our data demonstrate that one of the genes whose expression is strongly repressed by ZEB proteins is DNA ligase 1, *LIG1*. LIG1 catalyses joining Okazaki fragments during DNA replication, and is implicated in Nucleotide Excision and Homologous Recombination DNA repair pathways. Although functional redundancy in LIG1 and another DNA ligase, LIG3 has been reported, down-regulation of LIG1 during an EMT results in the delay in single strand breaks repair induced by low-dose of UV.

We show that ZEB proteins arrest cells in G1 phase of the cell cycle, and reduce transcription of *LIG1* in Rb-E2F-dependent manner. Cell cycle arrest largely occurs through up-regulation of the cyclin dependent kinases inhibitor p27Kip1. Knockdown of p27KIP1 uncouples EMT from cell cycle arrest and down-regulation of LIG1 resulting in LIG1-deficient actively proliferating cells. We show that these cells are characterised by a high level of DNA damage and elevated chromosomal aberrations. We also show that depletion of LIG1 is by itself sufficient for the induction of chromosomal instability.

EMT programs lead to the formation of a pool of dormant circulating tumour cells, which extravasate at distant organs and undergo mesenchymal epithelial transition (MET) to form metastases. We modelled a transient EMT in vitro, and found that cell cycle re-entry is mutagenic and associated with reduced LIG1 expression and enhanced level of fusions. Additionally, transient EMT results in a fluctuation in the average number of chromosomes and increases cell diversity with regard to the level of ploidy. We propose that transient EMT contribute to genomic instability in cancer.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my supervisors Dr Eugene Tulchinsky and Dr Steven Foster for their guidance, endless support, and encouragement throughout my PhD. Their enthusiasm, dedication and knowledge is the true inspiration that boosted my passion to scientific research. I would also like to thank my supervisor Professor G Don Jones for his support in the last year.

I would also like to thank all of the past and present members of the EMT group. A special thanks to Dr Marina Kriajevska for her invaluable support and advice with lab work. I would also like to extend my thanks to my fellow PhD students, Dr Youssef Alghamdi, Dr Gina Tse, Dr Qais Ismaeel, Zamzam Almutairi, Nabil Jaunbocus, Dr Hana Al-Mahmoodi and Dr Ban Alwash, who gave such a nice and exciting journey. Thanks for your friendship, encouragement and invaluable support throughout my PhD. A special thanks also to Dr Kees Straatman for his help with karyotyping microscopy. I would also like to thank my friends from Dr Steven Foster lab, Dr Kheloud Alhamoudi and Salwa Al-mayouf for their support, advice and being always beside me whenever I needed.

I would also like to thank all of my friends who supported me throughout my journey specially Dr Madina Hasan, Shashi Rana, Attiya Khan, Aisha Khan, Dr Heba Alhamal, Dr Marwah Aldriwesh and Rabab Al Sunni. I cannot thank you enough for being by my side through ups and downs and for giving me the most wonderful time outside the lab.

Mum and Dad! No words are enough to express how grateful I am to you. I cannot thank you enough for your unconditional love, support and encouragement throughout my life. Thanks for investing and believing on me and my siblings. You are my inspiration and without you by my side, honestly, I would not have come this far.

I would also like to thank my brothers Mustafa, Ahmed, Awad and Ali; and my sister Zahra for your unconditional support and love throughout my journey. Many thanks to the rest of my family for their continued prayers and wishes.

A very special thanks to my dear husband Zaky for supporting me through the extreme ups and downs throughout my PhD, especially during my thesis writing. Thanks for listening to my work description and analysis even though you di not understand the science. I would also like to thank my little boy Ahmed for giving me a reason to smile at the end of long tiring days.

Last but not the least, I would like to thank my sponsor, the ministry of Education for supporting me and giving me the opportunity to do my PhD.

Dedication

I would like to dedicate this thesis to my Aunts Fatima and Haya, who I lost few years ago due to cancer, and to whom I made a promise to do my best to help in fighting cancer. This is just a beginning! May Allah have mercy on you and I will always remember you for as long as I shall live.

Table of Contents

List of tables

List of Figures

Abbreviations list

4-OHT	(Z)-4-Hydroxytamoxifen
aa	Amino acid
ABC	ATP binding cassette
AdD	Adenylation domain
AMP	Adenosine 5'-monophosphate
AP	Apurinic/apyrimidinic
APS	Ammonium persulphate
ATLD	Ataxia telangiectasia (AT)-like disorder
ATM	Ataxia telangiectasia mutated
ATP	Adenosine Triphosphate
BER	Base excision repair
BN	Binucleated
BRCA	Breast cancer early onset protein
BRCT	BRCA1 carboxy terminal
BSA	Bovine serum albumin
CAK	CDK activating enzyme
CDK	Cyclin dependent kinase
CDK-I	Cyclin dependent kinase inhibitors
CHX	Cycloheximide
CID	CtBP interacting domain
CINK4	Cdk4/6 Inhibitor IV
CSC	Cancer stem cells

CtBP	C-terminus binding protein
CtIP	CtBP-interacting protein
DAPI	4′,6-diamidino-2-phenylindole
DBD	DNA binding domain
DDR	DNA damage response
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco’s Modified Eagle Media
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DOX	Doxycycline
DSB	Double strand break
DSBR	Double-strand break repair
E-cadherin	Epithelial cadherin
ECL	Enhanced chemiluminescence
ECM	Extra cellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EMT	Epithelial-to-mesenchymal transition
EMT-TF	EMT transcription factors
EtBr	Ethidium Bromide
EtOH	Ethanol
FA	Fanconi Anaemia
FACS	Fluorescence-activated cell sorting

FBS	Foetal bovine serum
HDAC	Histone deacetylase
HGF	Hepatocyte growth factor
HJs	Holiday junctions
HNPCC	Hereditary nonpolyposis colon cancer
HR	Homologous recombination
HRP	Horseradish peroxidase
IR	Ionizing radiation
kDa	kilo Dalton
KID	Kinase inhibitory domain
LIG1	DNA ligase I
LIG3	DNA ligase III
LIG4	DNA ligase IV
MDCK	Madin-Darby Canine Kidney
MeOH	Methanol
MET	Mesenchymal-to-epithelial transition
miRNA	microRNA
MMP	Matrix metalloproteinases
MMR	Mismatch repair
MN	Micronuclei
MRN	MRE11-RAD50-NBS1
MW	Mowat-Wilson
N-cadherin	Neural cadherin

NE	Nuclear envelope
NER	Nucleotide excision repair
NES	Nuclear export signal
NHEJ	Non-homologous end joining
NLS	Nuclear localisation sequence
NOD/SCID	non-obese diabetic/severe combined immunodeficient
NTase	Nucleotidyltransferases domain
OB	Oligonucleotide/oligosaccharide binding
ONPG	o-Nitrophenyl β-D-galactopyranoside
PBS	Phosphate buffered saline
PCAF	p300/CBP associating factor
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PenStrep	Penicillin Streptomycin
PGDF	Platelet derived growth factor
Pgp	P-glycoprotein
PI	Propidium iodide
PIP	PCNA interacting peptide
Rb	Retinoblastoma susceptibility protein
RFTS	Replication factory-targeting sequence
RNA	Ribonucleic acid
ROS	Reactive oxygen species
RPA	Replication protein A

RPMI	Roswell Park Memorial Institute medium or RPMI 1640
RS-SCID	Ionizing radiation sensitivity with severe combined immunodeficiency
RT	Room temperature
RT-qPCR	Real time quantitative reverse transcription PCR
SD	Standard deviation
SDS	Sodium dodecyl sulphate
SDSA	Synthesis-dependent strand annealing
siRNA	Small interfering RNA
SSB	Single strand break
ssDNA	Single stranded DNA
TAE	Tris-acetic acid-EDTA
TBS	Tris-buffered saline
TBST	Tris-buffered saline-Tween 20
TCR	Transcription coupled repair
TEMED	N,N,N’,N’-tetramethylethylenediamine
TGF-β	Transforming growth factor-β
UV	Ultraviolet
V(D)J	V (variable), D (diversity), and J (joining)
WT	Wild type
β-ME	2-mercaptoethnol

Chapter 1 : Introduction

1.1 The hallmarks of cancer

The term cancer describes the state of atypical and uncontrolled proliferation of cells. Cancer is believed to develop in a multistep process during which the cells acquire many genetic and epigenetic modifications in their oncogenes and tumour suppressor genes. These modifications as well as inflammation are thought to ultimately result in the acquisition of the cancer's hallmarks proposed by Hanahan and Weinberg (Hanahan and Weinberg, 2000, 2011). In the course of the multistep cancer development, the tumours are thought to acquire six distinctive and complementary capabilities that allow them to grow and eventually metastasize (Figure 1.1). The first capability is the maintenance of proliferative signalling, which can be evolved through various routes including growth signal self-production and autocrine stimulation. Additionally, proliferation can be maintained through the stimulation of the microenvironment cells to produce growth signals, cancer cells hyper-sensitisation to limited levels of ligands, the induction of ligand independent activation of growth receptors and the constitutive activation of downstream effectors of the growth receptors (Bhowmick et al., 2004; Witsch et al., 2010). The second hallmark is to escape growth suppression, which is usually accomplished through the inactivation of tumour suppressors, such as the retinoblastoma (Rb) protein, which functions as a gatekeeper protein for the G1 cell cycle checkpoint. The third capability is to resist apoptosis, which could be achieved through either dampening of pro-apoptotic sensors, upregulation of survival signals or anti-apoptotic proteins or by-passing apoptotic pathway activation through extrinsic ligand induction. The fourth capability is to achieve replicative immortality, which is usually linked to the maintenance of the length of telomeres, for instance, through telomerase upregulation (Blasco, 2005). The fifth of the hallmarks is to be able to induce the formation of new blood supply (angiogenesis), which usually happens via unbalanced pro-/ anti-angiogenic factors. The sixth is the ability to invade and metastasise, which usually involves the ability of the cancer cell to alter its shape and its adhesion to other cells or to the extracellular matrix (ECM) (Cavallaro and Christofori, 2004). These capabilities are acquired successively and differently by different tumours depending on the different types of cells that constitute the tumour. It has been suggested that the mere determination of tumour cells traits

would not be sufficient for the proper understanding of cancer biology. Alternatively, this has to be expanded into studying the tumour microenvironment or tumour associated cells, which are actively involved in tumourigenesis as well as the development of cancer hallmarks (Hanahan and Weinberg, 2011).

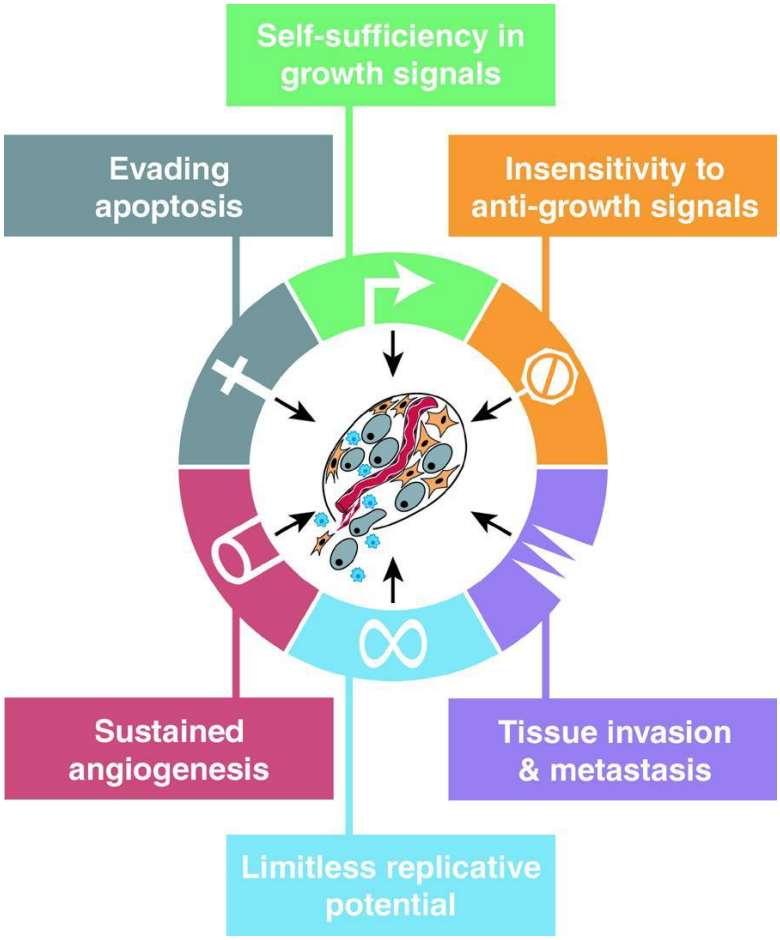


Figure 1.1: The original hallmarks of cancer.

An illustration of the original hallmarks of cancer proposed by Hanahan and Weinberg in 2000. Arising research has already improved our understanding of these capabilities and the mechanisms behind them. Image is taken from (Hanahan and Weinberg, 2000).

These comprise the original suggested hallmarks of neoplasia. As a result of ongoing research in the following decade, two more capabilities have been added to the list (Figure 1.2). The first of these hallmarks is to gain the ability to reprogram cellular metabolism in a manner that supports the growth and proliferation, survival and functionality of the various types of cells within the tumour. The second of the emerging hallmarks is the capability to escape destruction imposed by the various components of the immune system. This could infer a duality of function that can be taken by immune system cells, as

they can perform an enhancing or an antagonising roles on the development and progression of tumours (Hanahan and Weinberg, 2011).

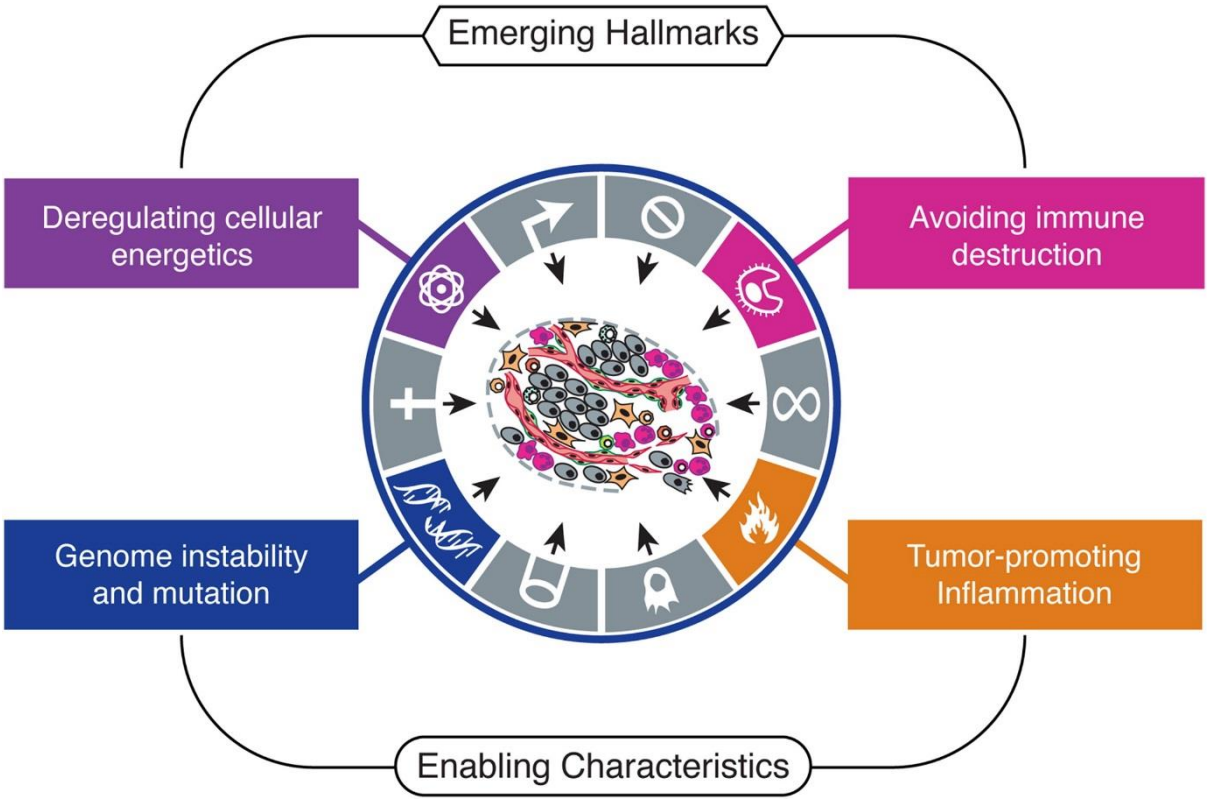


Figure 1.2: The emerging hallmarks and enabling features of cancer.

An updated illustration of the hallmarks of cancer with the new two emerging hallmarks as well as the enabling features proposed by Hanahan and Weinberg in 2011. Image is taken from (Hanahan and Weinberg, 2011).

In addition, another two characteristics were recognised recently that are thought to empower the original and emerging hallmarks in a manner that would enhance tumourigenesis (Figure 1.2). The first of these enabling features is the accumulation of mutations and genomic instability, which could be achieved through mutating or epigenetically inactivating tumour suppressors and the genes of DNA surveillance machinery. Examples of such genes include those implicated in DNA damage detection and subsequent repair activation, those with direct involvement in DNA repair pathways as well as the antagonists of mutagenic agents. Additionally, telomeres erosion adds another level of genomic instability (Artandi and DePinho, 2010). So, genomic instability would

give a selective advantage to the tumour cells against apoptosis or senescence conferred by such machineries and hence, their dominance and outgrowth and the subsequent formation of new tumours (Hanahan and Weinberg, 2011; Jackson and Bartek, 2009). As a result of the accumulating mutations, tumour progression generates new genomic makeup and hence, another level of complexity for metastatic cancer treatment. The second of the enabling characteristics is the acquisition of pro-tumourigenic inflammation. Previously, inflammation has been thought to be an attempt made by the immune system to eliminate a tumour. However, recent research has provided evidences of tumour promoting inflammation. This paradoxical effect has been demonstrated to be in large imposed by cells of the innate immune system. Such potentiation roles include enriching the tumour associated microenvironment with various factors required for growth, survival, angiogenesis induction, ECM modification, epithelial-to-mesenchymal transition (EMT) induction as well as invasion and metastasis. Additionally, immune system cells has also been suggested to release mutagenic chemicals that can in turn contribute in the genomic instability and foster the malignant features of cancer cells (Grivennikov 2010). Therefore, immune response could escalate the acquisition of the hallmarks (Hanahan and Weinberg, 2011).

1.2 Cancer prevalence

According to world health organisation (WHO), cancer is the second leading cause of death worldwide, where 1 in 6 deaths is because of cancer. The most common cancers globally are lung, liver, colorectal, stomach and breast, as they, together, form more than half of the cancer related deaths (WHO, 2018). Additionally, cancer is thought to be the cause for around 28% of deaths in the UK in 2016 with lung, bowel, breast and prostate cancers accounting for around 50% of all cancer related deaths. In the period of 2014-2035, it is estimated that cancer incidences for all types would increase by a range of around 2% to 70% for some forms of cancer (cancer research UK, 2016). Poor lifestyle, smoking and alcohol are believed to be the main reasons behind this increase in cancer incidences. The main causes of cancer related mortality are metastasis and therapeutic resistance (Seyfried and Huysentruyt, 2013; Talmadge and Fidler, 2010).

1.3 The invasion-metastasis cascade

The invasion-metastasis cascade refers to the multi-step process through which cells from a primary tumour acquire the ability to disseminate and eventually form a secondary tumour (Hanahan and Weinberg, 2011; Seyfried and Huysentruyt, 2013; Talmadge and Fidler, 2010). At the first stage of this cascade, some cells develop a capability to locally invade and then intravasate neighbouring blood/lymphatic vessels. This is then followed by the dissemination in the blood/lymphatic vessels before extravasation at the destined site. After that, those cells form small patches of tumours (micrometastases), which eventually colonise and form macroscopic lesions (Figure 1.3). This process is thought to be achieved through waves of the reversible processes of EMT and mesenchymal-to-epithelial transition (MET). It is believed that cells within the primary tumour receive signals to undergo an EMT (i.e. gain invasive capabilities) where they start to dissociate from the primary tumour and start intravasation and dissemination in the blood steam. As they travel far from the primary tumour site, the gradient of those signals decreases and the responsiveness of the tumour cells decline. As a result, the tumour cells might undergo MET and extravasate vasculature and start colonisation at the new (i.e. secondary) site, forming a metastasis with cells that are histopathologically resembling primary tumour cells (Géraud et al., 2014; Hugo et al., 2007).

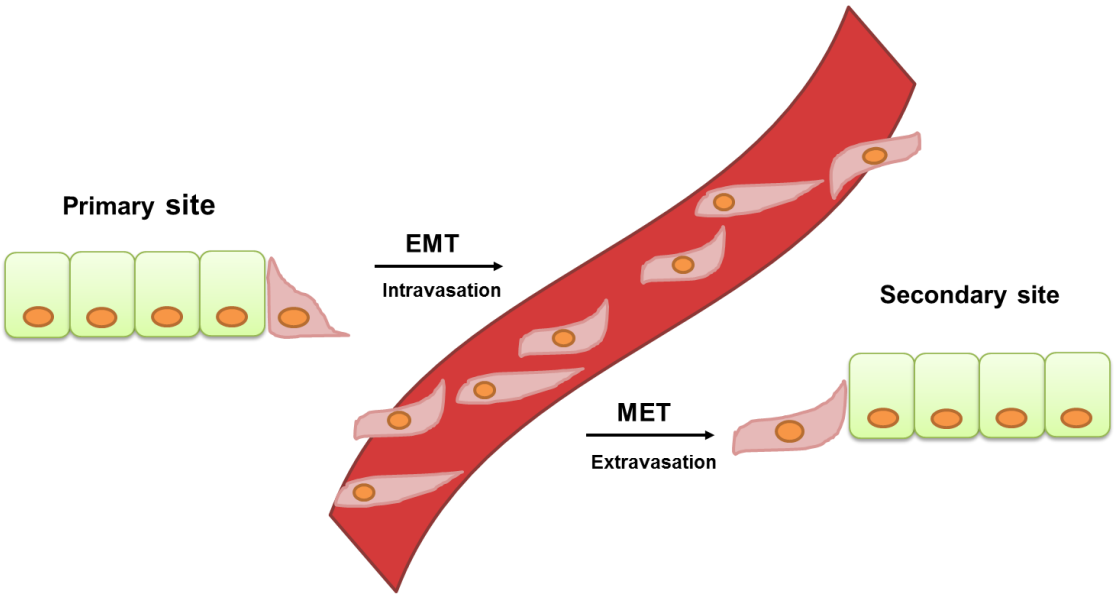


Figure 1.3: The invasion-metastasis cascade.
An illustration of the sequence of events of metastasis that enables carcinoma cells from primary tumour to disseminate and travel around the body to colonise a secondary tumour. Figure adapted from (Hilmarsdottir et al., 2014).

The choice of the secondary site at which the metastasis would form is dependent on the response of the recipient organ and the degree of congeniality between the tumour cell and the recipient organ (Talmadge and Fidler, 2010). For instance, studies have demonstrated that some hosts could produce anti-tumour signals, which restrict the survival and growth of the tumour cells. This could explain why not all the organs contain metastases and that different tumours metastasise to specific tissues. For example, breast cancer cells preferentially metastasise to the liver or bone marrow rather than the spleen (Géraud et al., 2014).

1.4 EMT and EMT-transcription factors (EMT-TFs)

EMT was initially defined as the transformation of epithelial cells to mesenchymal ones based on phenotypic alterations observed in the studies of the chick embryo primitive streak (Hay, 1995). However, over a decade ago, at the first meeting of the EMT international association, the term “transformation” was changed into the term “transition” reflecting a rather dynamic and highly plastic character of the process. EMT refers to the state at which a cell becomes more mesenchymal through the acquisition of migratory and invasive behaviour by modifying intercellular, and cell-ECM adhesion, and organisation of the cytoskeleton. In turn, MET occurs when cellular state shifts toward more epithelial appearance through the loss of migratory potential and the acquisition of hallmark epithelial features, which are the expression of the different adhesion molecules and apico-basal polarisation (Lamouille et al., 2014; Nieto et al., 2016; Thiery et al., 2009). Recent research, however, has modified our understanding of EMT/MET processes in cancer, and established that cells are actually moving through a spectrum of intermediate phases between the epithelial and mesenchymal states. This suggests that cells would not necessarily undergo a full EMT or MET, but instead they experience a partial transition, namely co-expression of epithelial and mesenchymal markers. Signals that tumour cells receive from the microenvironment, as well as the mutational status define to which extent epithelial or mesenchymal features are developed (Figure 1.4).

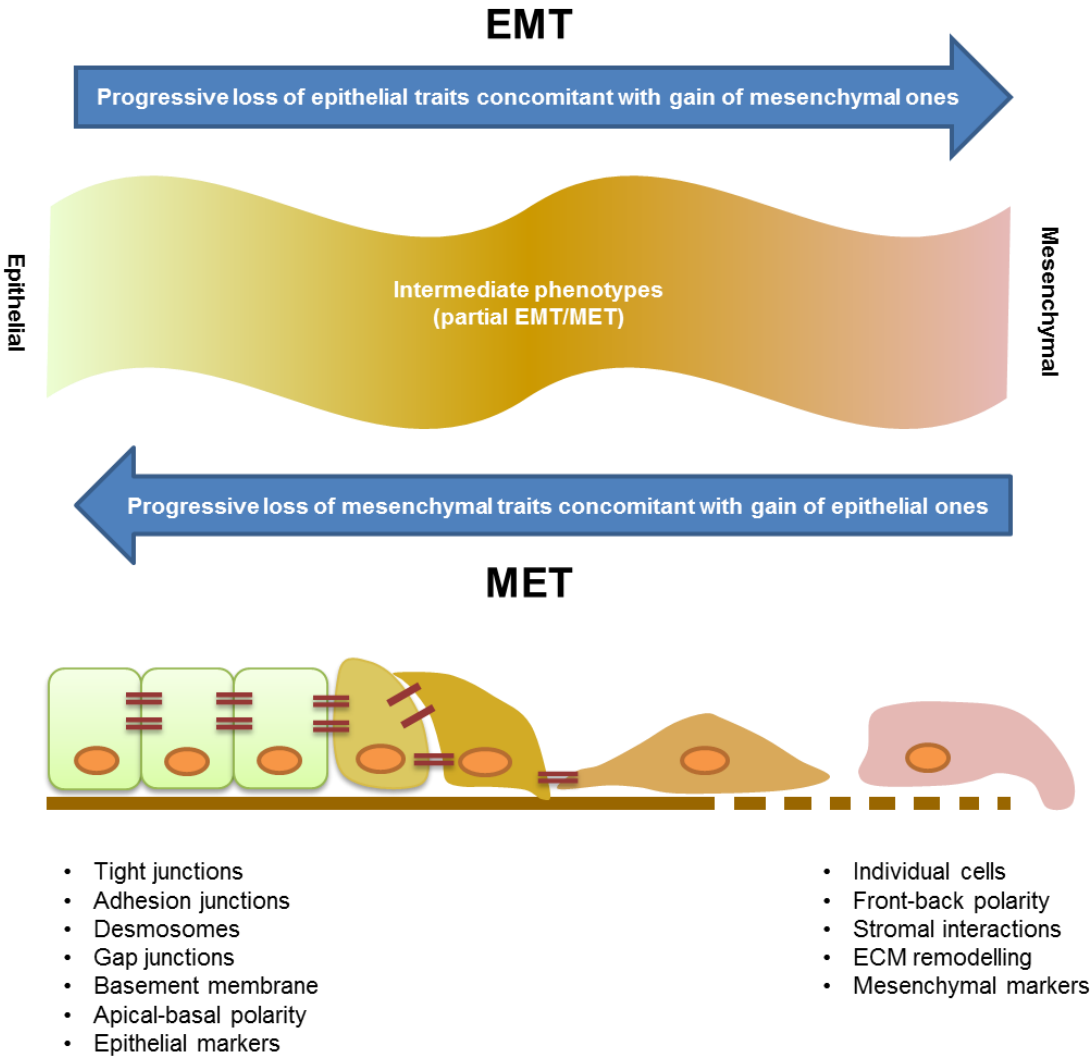


Figure 1.4: The plastic and dynamic nature of EMT/MET.

An illustration of the progressive changes in cell morphology and behaviour between the two sides of EMT/MET spectrum. At the end of the epithelial side, cells express the various forms of cell-cell and cell-ECM adhesion molecules, have an intact basement membrane and a well-defined apico-basal polarity. However, as they start transiting down the epithelial-mesenchymal spectrum they gradually lose cell contacts, remodel the ECM and acquire front-back polarity, but with residual epithelial features. Figure adapted from (Nieto et al., 2016).

The EMT is thought to be elicited as consequence of pleiotropic signals that are released from the different cells of the tumour microenvironment, including different growth factors and cytokines, such as transforming growth factor-beta (TGF-β), hepatocyte growth factor (HGF), epidermal growth factor (EGF), platelet derived growth factor (PGDF), etc. Those signals are thought to induce the expression of a group of transcription factors that cause EMT induction (EMT-TFs), several epigenetic and post translational modifiers and miRNAs (Marcucci et al., 2016).

All of these signals and subsequent EMT-TFs activation ultimately result in development of the EMT hallmarks. These hallmarks are epithelial adhesion molecules decomposition, apico-basal polarity loss. Those are concomitant with front-back polarity gain, cytoskeletal reorganisation resulting in morphology modification. Additionally, EMT induction involves repression of signature epithelial genes while activating mesenchymal ones, motility enhancement and the release of BM degrading enzymes which result in invasive potentials (Lamouille et al., 2014; Nieto et al., 2016; Puisieux et al., 2014; Tam and Weinberg, 2013). EMT-TFs directly repress transcription of genes encoding key epithelial markers that are involved in the formation of the tight junctions, adherens junctions, desmosmomes and gap junctions. Examples include zonula occludins 1 (ZO1, aka TJP1), epithelial cadherin (E-cadherin), desmocolins (DC1/2) and connexins, etc (Huang et al., 2012). In addition, EMT-TFs reduce or inhibit expression of the proteins involved in apical-basal polarity such as Crumbs complexes (CRB), which interacts with tight junctions at the apical compartment of the cell; and the Scribble complexes (SCRIB), which interacts with E-cadherin and delineate the basolateral compartment of the cell (Lamouille et al., 2014; Qin et al., 2005).

E-cadherin is a transmembrane cadherin junction protein that is expressed by epithelial cells, which is believed to serve a tumour and invasion suppressive function (van Roy and Berx, 2008). In fact, during multiple mechanisms of malignant transformation, E-cadherin has been demonstrated to be repressed and this downregulation has been linked to poor prognosis (Li et al., 2017b; Peinado et al., 2007). *CDH1* mutations have been reported in infiltrative lobular breast carcinomas as well as diffuse gastric carcinomas, which were thought to

produce a secreted protein with loss of its cell adhesion function (Berx et al., 1996). Although such mutations have been reported, most of the loss of E-cadherin functions in cancer has been linked to transcriptional repression of *CDH1* (hence, to EMT) (Vandewalle et al., 2009).

Simultaneously, EMT-TFs induce the expression of key mesenchymal markers that modify the cytoskeleton, cell-ECM interactions, and ensure mesenchymal type of cell-cell interactions. For instance, as cells down regulate E-cadherin, they upregulate the expression of the mesenchymal-type cadherin, neural cadherin (N-cadherin), hence, modifying the mode of cell adhesions. Involvement of N-cadherin in the interactions between mesenchymal cells facilitates their migration and promotes invasiveness. In addition, EMT-TFs alter the expression of intermediate filaments constituting proteins, which are essential cytoskeletal components for organelle trafficking within the cell. For instance, vimentin, which is a type III intermediate filament that is expressed in mesenchymal cells, is upregulated to replace cytokeratins (constituents of intermediate filament in epithelial cells). EMT-TFs also upregulate the expression of proteases, such as matrix metalloproteinases 2 and 9 (MMP2/9), which promote the degradation of the BM and ECM. All of these modifications induced by EMT-TFs in the course of EMT enhance the motile and invasive capabilities of cells (Lamouille et al., 2014; Thiery et al., 2009).

The process of EMT is implicated in normal physiological conditions such as wound healing in response to injuries, gastrulation, neural crest delamination and internal organs formation in embryonic development (Thiery et al., 2009). In pathological conditions such as malignant tumours and organ degeneration lesions like fibrosis, this process has been hijacked and exploited (Iwano et al., 2002; Puisieux et al., 2014). Studies have demonstrated that cellular plasticity during carcinoma development significantly mirrors embryonic cellular plasticity.

The process of EMT has been considered as a fundamental inducer of phenotypes associated with tumours that originate from epithelial tissues (Nieto et al., 2016). However, an increasing body of evidence showed that the activation of EMT-TFs in non-epithelial and mesenchymal tumours resulted in the activation of an EMT like process with a gain of mesenchymal and malignant properties.

Examples of such tumours include brain tumours, sarcomas and tumours of the haematopoietic system (Brabletz et al., 2018; Kahlert et al., 2017). For instance, ZEB1 protein expression in glioblastomas was found to induce neural stem cell markers and enhance their tumorgenicity and chemoresistance (Siebzehnrubl et al., 2013). Additionally, TWIST1 activation in paediatric anaplastic large cell lymphoma (ALCL) associates with invasiveness and therapy resistance (Zhang et al., 2012). The association between EMT/EMT-TFs various processes will be discussed subsequently, which emphasise the significance of the continued research as a fundamental tool of tumour malignancies.

Examples of EMT-TFs include members of the TWIST bHLH proteins, TWIST1 and 2, the SNAIL family of the zinc finger proteins and members of the ZEB family of zing finger and E-box binding proteins (e.g. ZEB1 and 2, aka TCF8 and SIP1, respectively). For EMT progression, they usually regulate expression of each other and control transcription of target genes in a cooperative fashion. EMT-TFs are grouped based on the mechanism through which they repress the transcription of the *CDH1* gene, which codes for E-cadherin. Some of them repress the *CDH1* promoter directly (e.g. TWIST1, SNAIL, ZEB, E47 and KLF8), while others (Goosecoid, E2.2 and FoxC2) repression is indirect (De Craene and Berx, 2013; Lamouille et al., 2014; Thiery et al., 2009; Yang et al., 2010). As aforementioned, loss of E-cadherin expression is considered a hallmark of EMT induction in carcinomas as this would weaken stability of the adherens junctions.

EMT in cancer was initially described in cell cultures. However, EMT role in vivo is widely reported. For instance, mesenchymal circulating tumour cells (CTCs) isolated from the blood of women with metastatic breast cancer were found to express key regulators of EMT, including components of TGF-β signalling pathway and the transcription factor FOXC1. Serial monitoring of these cells showed a correlation with disease progression. This gave an evidence for EMT association with dissemination of human cancer cells (Yu et al., 2013). Additionally, Markiewicz et al. (2019) recently demonstrated that mesenchymal or partially mesenchymal CTCs isolated from metastatic breast cancer disease patients express EMT markers and correlate with aggressive phenotype, higher risk of death and higher risk of metastasis to lymph nodes. Moreover, a study comparing the molecular signature of tumour budding and tumour bulk cells

isolated from patients with colorectal cancer showed an EMT-positive signature in the cells of the tumour budding region. Budding cells also showed a phenotypic switch and acquired migratory capability concomitant with proliferation reduction (De Smedt et al., 2017).

The basic helix-loop-helix (bHLH) family of EMT-TFs

The superfamily of bHLH transcription factors are homo- and heterodimeric transcription regulators which play a role in cell differentiation and lineage specification. Examples of the bHLH transcription factors include the inhibitor of differentiation (ID), E12, E47, TWIST1 and TWIST2, which are also fundamental regulators of EMT progression. They bind promoter DNA at the E-box consensus region (CANNTG) (De Craene and Berx, 2013; Lamouille et al., 2014; Peinado et al., 2007). Like other EMT-TFs, upon their expression, bHLH TFs downregulate *CDH1* and upregulate *CDH2* (N-cadherin encoding gene) expression.

TWIST expression during development as well as during cancer progression is triggered by several signalling pathways and cellular conditions. For example, in the hypoxic conditions, the hypoxia-inducible factor 1α (HIF1α) transcription factor promotes EMT via TWIST1. Additionally, TGF-β induced repression of ID proteins, which associate with and inhibit the function of TWIST1 or other bHLH factors, induces the expression of TWIST1 and downstream EMT. Stability of the TWIST1 protein is controlled by ubiquitination-proteasome pathway. Oncogenic pathways induce EMT by counteracting degradation of TWIST1 by proteasome. For example, TWIST1 is protected from ubiquitination through phosphorylation at Ser68 by MAPKs, which thereby promote its nuclear localisation and function (Lamouille et al., 2014).

TWIST is upregulated in various cancers and linked to invasiveness, metastasis, fatal clinical outcome and poor survival. For instance, elevated levels of TWIST expression in ductal carcinoma was found in association with high-grade tumours (Mironchik et al., 2005). Additionally, high levels of TWIST were implicated in promoting the metastasis of other solid tumour types, such as oesopharengeal squamous cell carcinoma and prostate cancer (Kwok et al., 2005; Peinado et al., 2007; Yuen et al., 2007).

SNAIL superfamily of EMT-TFs

In vertebrates, the SNAIL superfamily of transcription factors comprises three members SNAIL1, SNAIL2 and SNAIL3 (aka SNAIL, SLUG and SMUC, respectively). Two of these are implicated in EMT induction, which are SNAIL1 and SNAIL2. They repress E-cadherin by directly binding the consensus E2-box sequence (C/A (CAGGTG)), within the *CDH1* promoter proximal region, with their carboxy-terminal zinc-finger domains (Lamouille et al., 2014; Nieto, 2002; Peinado et al., 2007). SNAIL proteins recruit the Polycomb repressive complex 2 (PRC2) for histone modifications (Herranz et al., 2008). The PRC2 complex consists of methyl transferases, the enhancer of zeste homologue 2 (EZH2), G9a and suppressor of variegation 3-9 homologue 1 (SUV39H1) (Dong et al., 2012a; Dong et al., 2012b). Interactions between SNAIL proteins, the transcriptional co-repressor SIN3A, histone deacetylases 1, 2 and/or 3, and the lysine-specific demethylase 1 (LSD1) are required for transcriptional repression of *CDH1* (Lin et al., 2010). These PRC2 subunits are implicated in H3 histones methylation and acetylation of the lysine residues at positions 4 (H3K4), 9 (H3K9) and 27 (H3K27). This illustrates the repressive mode of action for SNAIL-TFs, but they simultaneously activate mesenchymal genes (e.g. *CDH2*) (Bernstein et al., 2006). Their dual role in transcriptional regulation is explained by their potential to cause repressive trimethylation at K9 and activating acetylation at K18 on Histone 3. At many target promoters of embryonic stem cells, SNAIL proteins generate the so called bivalent domains to which both activating and repressing marks can be attached. This is thought to maintain the promoter at a balanced state where it would be repressed in the absence of the differentiation signals but also enable temporal activation once such signal is received (Bernstein et al., 2006; Lamouille et al., 2014).

SNAIL expression, and hence EMT induction and progression, is activated by the cooperation of multiple signalling pathways, including TGF-β, WNT proteins, Notch and GFs. After their induction, SNAIL proteins are thought to control the expression of their target genes by interacting with other TFs (Lamouille et al., 2014). For instance, for the activation of MMPs, SNAIL1 cooperates with MAPK-activated TF ETS1 (Jorda et al., 2005). Likewise, it interacts with SMADs 3 and 4 to repress genes coding for E-cadherin and occludin following TGF-β induced

EMT (Vincent et al., 2009). SNAIL protein stability is controlled by phosphorylation and nuclear export, followed by ubiquitination and degradation. For instance, SNAIL1 phosphorylation at serines 79 and 101 within the first serine rich-motif, which is mediated by the glycogen synthase kinase-3β (GSK-3β) exports it from the nucleus. After that, SNAIL1 is phosphorylated again at serines 108, 112, 116 and 120 of the second serine-rich motif, which enables its degradation by ubiquitin-proteasome pathway. All of these destabilising signals are targets for inhibition by EMT initiating signals (Lamouille et al., 2014).

SNAIL proteins play crucial roles in multiple developmental processes that involve EMT. Indeed, SNAIL protein expression is vital for the formation of the mesoderm, gastrulation, and delamination of the neural crest cells. Mouse embryos that lack SNAIL where found to have defective mesodermal cells that lack the migratory potential and are unable to downregulate E-cadherin (Carver et al., 2001).

SNAIL proteins have been implicated in several invasive carcinomas (Peinado et al., 2007), their presence in tumour cells is linked to metastasis, tumour relapse and poor prognosis. For instance, SNAIL1 expression has been shown to mediate the repression of E-cadherin and metastasis and inversely correlate with tumour cells differentiation grade in breast carcinoma (Blanco et al., 2002). Similarly, SNAIL2 expression in colorectal cancer and melanoma correlates with poor prognosis and patient survival and malignancy, respectively (Gupta et al., 2005; Shioiri et al., 2006).

Zinc finger and E-box binding homeobox (ZEB) family of EMT-TFs

In vertebrates, the ZEB family of transcription factors is represented by two members, ZEB1 and ZEB2 (aka TCF8/δEF1 and SIP1, respectively), that are encoded by two different genes (*ZFHX1A* and *ZFHX1B*, respectively). The ZEB TFs are characterised by the presence of two clusters of zinc-finger DNA binding domains (ZFD) at both of the N- and the C-termini with a homedomain (HD) localised between two ZFD clusters (Figure 1.5) (Funahashi et al., 1993; Verschueren et al., 1999). ZFDs are highly similar between the two proteins, however, the sequence outside is less conserved. The HDs of ZEB1 and ZEB2 are believed to participate in protein-protein interaction rather than DNA binding. At both sides of the HD and between the two ZFD clusters lie the binding domains involved in the interactions with the factors regulating functions of ZEB TFs (Vandewalle et al., 2009; Zhang et al., 2015). Examples of such domains in ZEB include the CTBP-interaction domain (CID); SMAD-binding domain (SBD); NuRD-binding domain characterised at the N-terminus of ZEB2. ZEB proteins are thought to act as both transcriptional activators and repressors dependent on the cell type and DNA context (Postigo et al., 2003; Remacle et al., 1999).

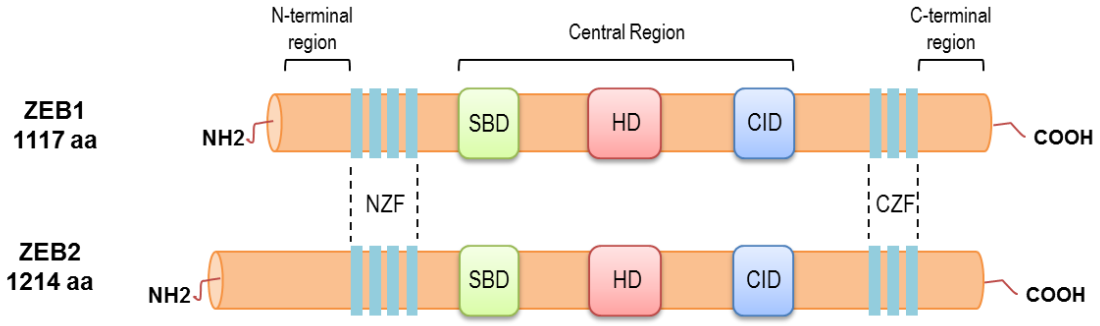


Figure 1.5: A scheme of ZEB proteins.

ZEB proteins have two clusters of zinc finger binding domains at the N-terminal (NZF) and the C-terminal (CZF). In addition, they contain a central region that is composed of a Smad-interacting domain (SBD), a central homeodomain (HD) and a CtBP-interacting domain (CID). Figure modified from (Vandewalle et al., 2009)

ZEB proteins function by concomitant binding to the two regulatory E-boxes (CACCT and CACCTG) within the promoters of their target genes (De Craene and Berx, 2013; Lamouille et al., 2014; Peinado et al., 2007; Vandewalle et al.,

2009). The position and the distance between the two E-boxes can change in different target promoters. To achieve transcriptional repression, ZEB proteins usually recruit a transcriptional co-repressor such as the C-terminal binding protein (CTBP) (Postigo and Dean, 1999). However, CTBP independent repression can occur with the aid of chromatin modifying proteins (Peinado et al., 2007). For instance, ZEB1 has been shown to repress *CDH1* transcription by recruiting BRG1, which is the ATPase subunit of the chromatin remodelling complex SWItch/Sucrose Non-Fermentable (SWI/SNF) (Reisman et al., 2009; Sanchez-Tillo et al., 2010). To activate transcription, ZEB proteins recruit transcriptional co-activators such as p300/CBP-associated factor (PCAF, aka KAT2B). Additionally, ZEB proteins can activate transcription through the histone demethylation by recruiting demethylases, such as the ZEB1-interacting lys-specific demethylase 1 (LSD1) (Peinado et al., 2007; Wang et al., 2007).

ZEB proteins are important players in TGF-β and BMP signalling, due to the presence of the SBD that enables their interaction with receptor activated Smads (Verschueren et al., 1999). ZEB1 and ZEB2 were suggested to play opposing roles in transcriptional regulation of TGF-β and BMP target genes (Postigo, 2003). For instance, downstream of TGF- β, ZEB1 has been shown to act as a transcriptional activator of the Smad-regulated genes through the recruitment of the transcriptional co-activators P/CAF and p300. On the other hand, ZEB2 mediates repression of TGF-β/Smad target genes through the recruitment of the CTBP co-repressor to the Smads (Vandewalle et al., 2009).

The expression of ZEB proteins is induced by several molecular pathways including TGF-β, WNT, nuclear factor-kappa B (NF-κB) as well as RAS-MAPK signalling activating by various growth factors (Figure 1.6) (Thiery et al., 2009). For example, TGF-β signalling induces the expression of the TF Ets1 that in turn induces the expression of both ZEB1 and ZEB2. Additionally, TGF-β signalling has been shown to repress the expression of miRNAs, which are known to degrade the mRNA of both of ZEB1/ZEB2 (Lamouille et al., 2014; Vandewalle et al., 2009). ZEB protein expression is induced by other EMT-TFs. For instance, ZEB1 expression is induced directly by SNAIL alone and the association between SNAIL and TWIST (Dave et al., 2011).

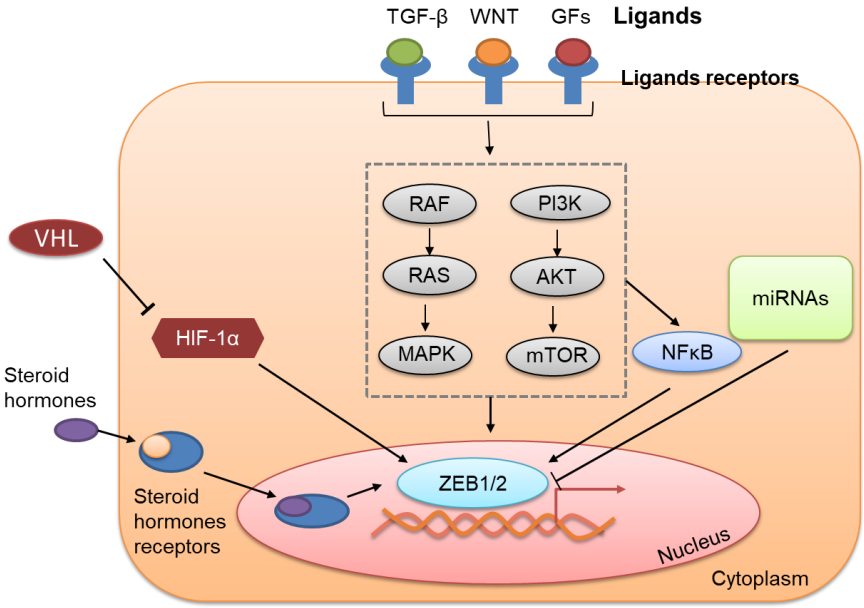


Figure 1.6: A simplified scheme of the upstream signalling events of the ZEB family of transcription factors.

The ZEB proteins are activated by downstream effectors of multiple signalling pathways such as TGF-β, NFκB, WNT, miRNA and Hypoxia and the HIF1-α signaling pathways. Another pathway of induction is through the MAPK pathway downstream of growth factors (GFs) activation. ZEB1 could also be activated downstream of steroid hormones receptors. The ZEB proteins can also be downregulated by TGF-β, like ZEB2 downregulation as a result of miR192 action. The figure is adapted from (Lamouille et al., 2014; Marcucci et al., 2016; Vandewalle et al., 2009).

ZEB proteins mRNA is degraded by miRNAs, such as members of the miR200 family as well as miR-205 (Hill et al., 2013). However, transcription of these miRNAs is also repressed by the ZEB proteins, hence generating a double feedback loop control. Additionally, ZEB proteins are postranscriptionally regulated by SUMOylation, which is the process through which a SUMO group (small ubiquitin like modifiers) is attached to the protein affecting protein structure and subcellular localisation (Lamouille et al., 2014; Vandewalle et al., 2009). ZEB2, for example, is SUMOylated by PRC2, which interferes with ZEB2/CTBP interaction, as well as induces its nuclear export, hence, blocking its repressive potential (Long et al., 2005).

In embryogenesis, ZEB proteins are key to the development of skeletal muscles, cartilage, bones, the heart, the central and peripheral nervous systems (CNS/PNS) and haematopoietic cells (Peinado et al., 2007; Vandewalle et al., 2009). For example, ZEB1 is thought to act as a positive regulator of neuronal differentiation by repressing transcription of the *RCOR1* gene. *RCOR1* encodes

RE1-silencing transcription factor (REST), which is a vital subunit of the BRAF-HDAC (BHC) complex, a repressor of the neuronal lineage (Ballas et al., 2001). ZEB1 expression has been detected in the notochord, somites, limb, derivatives of the neural crest and specific areas of the CNS (Darling et al., 2003). ZEB2 expression was detected in various parts of the peripheral nervous system and central nervous system (PNS and CNS) at early stages of developing mice and humans. Additionally, in adult mice and human foetuses (25-week old), ZEB2 expression can be detected in all neuronal regions, as well as several other tissues such as heart, liver, thymus, skeletal muscles, kidneys and bladder, which implies an important role for ZEB2 in the functions of these tissues (Bassez et al., 2004).

Zeb1-deficient mice develop *in utero* normally, but die after birth suffering from respiratory failure, several skeletal malformations like malformed limbs and sternum and craniofacial defects as well as severe T cell deficiencies (Takagi et al., 1998). All of these defects were shown to be linked to EMT as *Zeb1* loss was accompanied by E-cadherin re-expression and vimentin loss in multiple embryonic tissues (Liu et al., 2008b). However, the CNS of these mice was not found to harbour significant phenotypic defects, which suggest a possible compensation by *Zeb2*. *Zeb2* deficiency was shown to produce defects in mice after day 8.5 of development, where embryos exhibited failure of closure of the neural tube, loss of a proper neural plate/ectoderm boundary, short somites and loss of the first branchial arch. These embryos also displayed abnormal migration of neural crest cells that resulted in craniofacial, melanocytic and cardiogenesis defects as wells as multiple malformations in peripheral nervous system (PNS). The growth of these mice was severely abnormal, and they die by day 9.5 of development (Van de Putte et al., 2007; Van de Putte et al., 2003). Compound *Zeb1* and *Zeb2* deficiency produced defects that were reminiscent of *Zeb2* knockout but with worse neural tube closure failure and thinning of the dorsal neuronal tube portion (Vandewalle et al., 2009).

ZEB protein mutations that result in loss of their function has been linked to the diseases in humans, which highlight their importance for normal human development. Heterozygous mutations in *ZFHX1B,* for instance, which resulted in the generation of a truncated protein and loss of ZEB2 function were linked to

the Mowat-Wilson Syndrome (MWS) associated with the congenital mental retardation (Wakamatsu et al., 2001; Yamada et al., 2001). Patients with MWS have variable characteristic defects in the CNS (including mental retardation, hippocampal malformation, seizures and corpus callosum agenesis), as well as abnormalities that result from neural crest cells developmental abnormalities (such as aberrant facial features, heart defects and Hirschsprung disease) . Hirschsprung disease (aka aganglionic megacolon) is a disease that is characterised by the loss of neuronal ganglion cells in the colon (Mowat et al., 1998; Mowat et al., 2003; Vandewalle et al., 2009). Nonsense mutations and heterozygous frameshift of *ZFHX1A* were shown to cause posterior polymorphous corneal dystrophy (PPCD). PPCD is a rare autosomal dominant disease of the cornea that is characterised by overgrowth and transformation of the corneal endothelium to epithelial cells with an abnormally thickened basement membrane (Descemet membrane) with a lot of collagen aggregates in the cornea (Krafchak et al., 2005).

ZEB proteins are implicated in the progression of malignant cancers. This is usually associated with their ability to downregulate the expression of E-cadherin and other cell adhesion and cell polarity proteins as well as the induction of EMT, which is usually observed in advanced and metastasised disease. Abnormal expression of ZEB1 has been detected in multiple human cancers, including liver cancer, lung cancer, osteosarcoma, colon cancer, uterine cancer and breast cancer (Vandewalle et al., 2009; Zhang et al., 2015). ZEB1 was also shown to promote pancreatic cancer tumourigencity through the repression of miRNAs that inhibit stemness such as the family of miR200s and miR203 (Wellner et al., 2009). ZEB2 upregulation has been detected in various human cancer biopsies and was shown to correlate with poor prognosis, such as oral squamous cell, gastric, ovarian and breast carcinomas (Vandewalle et al., 2009). ZEB proteins have also been linked to senescence and therapeutics resistance that is linked to EMT. These features are to be discussed further later.

1.4.1 EMT subtypes

Depending on biological contexts and functional consequences, EMT has been divided into three different subtypes (Reviwed in (Kalluri, 2009) and (Kalluri and Weinberg, 2009)). Type 1 denotes the form of EMT that generates multiple types of cells with shared mesenchymal phenotypes without fibrotic or invasive characters. This form is implicated in implantation, embryogenesis and organogenesis. Examples of type 1 EMT include the generation of three germ layers (ectoderm, mesoderm and endoderm) from the zygote during gastrulation as well as neural crest delamination from the neuroectoderm.

The second type denotes the type of EMT that is activated following trauma and inflammatory injury and is associated with the generation of fibroblasts and other related cells that are important in tissue repair. This is thought to be involved in tissue reconstruction that takes place during regeneration, wound healing, as well as organ fibrosis. A fundamental difference between type 2 and type 1 EMTs is the involvement of an inflammatory response, as inflammation cessation weakened wound healing and tissue regeneration. This can portray the pathogenesis of organ fibrosis as an abnormally persistent inflammation which results in prolonged wound healing and tissue distraction.

The third type is the form of EMT that is associated with tumourigenesis and metastasis. Type 3 EMT occurs in carcinoma cells at the invasive front of primary tumours, which have acquired genetic and epigenetic modifications that have enabled them to gain proliferative as well as invasive potentials. Studies have shown that most of these modifications occur in the oncogenes and tumour suppressors, which sensitises these cells to EMT-inducing factors released by contiguous microenvironment. This results in physiological effects that are very different from those induced by the other two types. It is believed that the cells undergoing type 3 EMT are the ones that are competent to start a new invasion-metastasis cascade (section 1.3).

1.4.2 EMT-TFs and cancer stem cells

For many human cancers, a small population of cells combines the properties of malignant and stem cells. This population was termed tumour initiating cells (TIC) or cancer stem cells (CSC). These cells are able to divide asymmetrically, to self-

renew, and to produce more differentiated progeny; they exhibit resistance to apoptosis and the capability to initiate macroscopic metastasis (Cai et al., 2018; Talmadge and Fidler, 2010). CSC confer cancer heterogeneity, plasticity and resistance to chemotherapy and radiation. Additionally, it is believed that CSCs are the cells responsible for relapse following excision of a primary tumour and therapy.

The first demonstration of CSC was reported by Bonnet and Dick in 1997 in acute myeloid leukaemia (AML). They showed how AML was generated from progenitor cancer cells (i.e. CSC) using haematopoietic stem cells cellular hierarchy. They showed that these CSC, (or the SCID leukemia-initiating cells as they called them) within all subtypes of AML they analysed, and irrespective of their mature phenotype, were $CD34^{++}/CD38^{-}$, which is a characteristic phenotype of normal haematopoietic stem cells (Bonnet and Dick, 1997). Another demonstration of CSC came from a study in which subpopulations of tumourigenic cells of the $CD44^{+}/CD24^{-}$ phenotype were isolated from 8/9 patients with breast cancer, and then xenografted into mammary fat pad of NOD/SCID mice. These cells were highly tumourigenic, they were able to self-renew (i.e. generate a new $CD44^{+}/CD24^{-}$ subpopulation) and generate heterogeneous low-tumourigenic population of cells that resembled the primary tumour (Al-Hajj et al., 2003).

EMT-TFs plays a fundamental role in the generation of CSC and it endows them with therapeutic resistance. Mani et al (2008) has shown that the number of cells with self-renewing capabilities was elevated following a constitutive EMT induced by SNAIL, TWIST or TGF-β in immortalized human mammary epithelial cells (HMLEs). These cells were found to express markers of mammary epithelial stem cells, $CD44^{+}/CD24^{-}$. These cells were also shown to have the ability to form mammospheres in culture that were able to give rise to a full mammary epithelial tree following mammary fat pads implantation. In fact, this self-renewing ability and enhanced tumourigenecity was also observed following the induction of a transient EMT (SNAIL-ER or TWIST-ER). Additionally, if these cells were to be subjected to multiple cycles of mammosphere assays, newly formed cells were also found to have same self-renewing ability and tumourigenecity capabilities,

even in the absence of EMT. So, transient EMT provided the cells with heritable CSC features.

It has been suggested that generation of CSC via EMT-TFs occurs through exploiting multiple signalling pathways that are implicated in normal functions of stem cells. Examples of such signalling pathways include the TGF-β, WNT, sonic Hedgehog (Shh) and others (Cai et al., 2018). Generation of CSC was also achieved in experiments involving manipulation of the expression of other factors such as miRNAs and the components of tumour microenvironment. For instance, recent studies have shown that in hepatocellular carcinoma, TGF-β released by tumour associated macrophages (TAMs) induced an EMT, resulting in the generation of cells with stem-like properties (Fan et al., 2014). Additionally, as aforementioned, ZEB1 has been implicated in promoting stem-like properties in pancreatic and colon cancer and that ZEB1 knockdown resulted in the loss of these properties. This is thought to be achieved through the direct transcriptional repression of key stemness-inhibiting miRNAs in pancreatic cancer, such as miR200c, miR203 and miR183. As a consequence, the expression of stem cell factors such as Sox2 and the polycomb repressor Bmi1 were elevated in ZEB1-positive cells (Wellner et al., 2009).

1.4.3 EMT-TFs and therapy resistance

Tumour resistance to therapeutics (i.e. chemotherapy, radiation or targeted therapy) is one of the major causes of patients' death. This resistance usually develops due to genetic or epigenetic alterations within tumour cells or their microenvironment. This could either be inherent, where resistance-inducing factors are present within a proportion of tumour cells before the therapy even starts, or acquired, where these alterations are generated as a protective mechanism after the treatment has been initiated (Zahreddine and Borden, 2013). Unluckily, it is thought that the likelihood of resistance development in patients during the treatment period is quite high. For instance, the development of chemoresistant adenocarcinoma after surgery and adjuvant therapy occurs in 50-70% of adenocarcinoma patients (Castells et al., 2012). Moreover, around 20% of acute lymphoblastic leukaemia adult patients are diagnosed with inherent treatment resistance (Zahreddine and Borden, 2013). For this reason, it is very

important to elucidate the mechanisms through which cancer cells resist treatment, as in most of the cases, this is still to be elucidated.

Therapy resistance could arise through several ways such as modifying drug uptake or efflux, mutations or amplifications of drug targets, mutations that inhibit apoptosis and induce cell survival. Resistance could arise as a result of the inherent heterogeneity within the tumour, where a subpopulation could gain or already carry advantageous resistant mutations enabling them to expand during treatment. Alterations in the drug transport and metabolism are among the best studied mechanisms of drug resistance. They could involve the evolution of different modes of drug uptake or efflux, or mechanisms of drug detoxification. For example, some drugs could exert their effects by interacting with a membrane bound receptor, whereas others might require entering the cell to function (Gottesman, 2002; Zahreddine and Borden, 2013). A frequently observed modification leading to an increased efflux of drugs and other toxins is upregulated expression of membrane transporters. An example of these is enhanced expression of the membrane transporters ATP binding cassette (ABC) (Gottesman, 2002). Three members of the ABC family of transporters has been implicated in cancers resistance to several drugs, which are the permeability glycoprotein (P-gp, aka *MDR1* gene product, and ABCB1), Multidrug resistance-associated protein 1 (MRP1) and mitxantrone resistance protein (aka breast cancer resistance protein (BCRP) and ABC protein of the placenta (ABC-P)) (reviewed in Ambudkar et al. (2002); Fletcher et al. (2010)). Multiple hydrophobic drugs such as doxorubicin, for instance, are transported by P-gp. Resistance to doxorubicin has been linked to the upregulation of P-gp in several cancer types including multiple GI tract cancers, haematopoietic cancers, breast and ovarian cancers (Goldstein et al., 1989; Gottesman, 2002; Mechetner et al., 1998).

EMT-TFs induces therapy resistance in many cancer types, including chemoresistance, radioresistance and the resistance to a variety of targeted therapies, such as inhibitors of mutant EGFR (Zhang et al., 2015). For example, Saxena et al. (2011) has shown that the expression of the EMT-TFs TWIST, SNAIL or FOXC2 and the consequent EMT induction in immortalized and non-invasive cell lines resulted in an elevated expression of the ABC transporters, migration and invasion and drug resistance. Additionally, ZEB1 expression has

been shown to induce radioresistance in the radiosensitive breast cancer cell lines HMLE and MCF7 (Zhang et al., 2014b). Accordingly, ZEB1 knockdown in cell lines with a known radioresistent phenotype, such as the breast cancer cell line SUM159, resulted in radiosensitisation.

Moreover, in different contexts, therapy resistance has been linked to a role of EMT in the generation of CSCs. For multiple cancers, CSC has been suggested to contribute to the resistance to therapy. This is achieved through providing protection mechanisms against DNA damage such as the removal of reactive oxygen species (ROS), the advantageous activation of checkpoint responses to DNA damage and the subsequent induction of DNA repair. Moreover, CSC trigger cell survival through the induction of several survival signalling pathways such as NF-κB or NOTCH (Holohan et al., 2013). For example, ZEB1 expression in the gemcitabine-sensitive differentiated pancreatic cancer cell line BXPC3 induces an EMT, stemness and drug resistance (Wellner et al., 2009).

1.4.4 EMT-TFs and the cell cycle

Previous studies have shown that EMT-TFs can control cell cycle progression in a cell type-dependent manner. EMT-TFs arrest cell cycle progression in various human cancer cells. Vega et al. (2004) have shown that Madin-Darby Canine Kidney (MDCK) cells transfected with Snail were unable to progress through the Rb restriction point, hence, arrested in the G1 phase of the cell cycle. They demonstrated that this arrest was a result of a very strong down-regulation in *CCND2* expression, the gene that codes for the cell cycle regulator cyclin D2. They have shown that Snail directly represses *CCND2* through binding to two E-box consensus sites in the promoter region. Additionally, Snail-expressing cells were found to have high levels of the cyclin-dependent kinase inhibitor (CDK-I) p21. Moreover, the same cell cycle arrest was observed in squamous carcinoma cell line A431 with the doxycycline-regulated expression of ZEB2. This was due to the direct repression of *CCND1* transcription, which encodes for the cell cycle regulator cyclin D1, by ZEB2 through direct interaction with Z-boxes 1-3 in the *CCND1* promoter. It was also shown that cyclin D1 ectopic expression coupled with ZEB2 expression in these cells resulted in arrest release but did not interfere with EMT induction (Mejlvang et al., 2007). ZEB2 expression in bladder carcinoma cells was also shown to elevate G1 phase cells (Sayan et al., 2009).

Additionally, the expression of either *Twist* or *Snail* in kidney tubular epithelial cells (TECs) has been shown to arrest the cell cycle at G2. This arrest has been linked to the upregulation of p21 expression. This arrest contributed to kidney fibrosis by impairing cellular ability to proliferate and regenerate (Lovisa et al., 2015).

In other cellular contexts EMT may produce an opposite, stimulatory effects on cell cycle progression. For example, Snail expression in the mouse epidermis resulted in an elevation of cellular proliferation (Jamora et al., 2005). In addition, *Zeb2* expression is important for normal development of the hippocampus and the dentate gyrus in the mouse (Miquelajauregui et al., 2007).

1.4.4.1 The mammalian cell cycle

In mammalian cells, the cell cycle is a complex network of events that culminate with the division of a cell and the production of two new daughter cells. It consists of 4 phases, an initial gap phase (G1) during which a cell grows and prepares for the next phase (S), where the DNA is replicated, followed by another phase of growth (G2) in which a cell prepares for the last mitosis phase (M). The three initial phases (G1, S, and G2) are known as the interphase of the cell cycle. Whenever a cell is not proliferating, either when growth signals are absent or in the presence of DNA repair (checkpoints) signals, then it is in the quiescent state (G0). The transition between these phases is controlled by holoenzymes that are formed by coupling cyclin-dependant kinases (CDKs) and the cyclins (Figure 1.7) (Barbash and Alan Diehl, 2008; Casimiro et al., 2014; Malumbres and Barbacid, 2009).

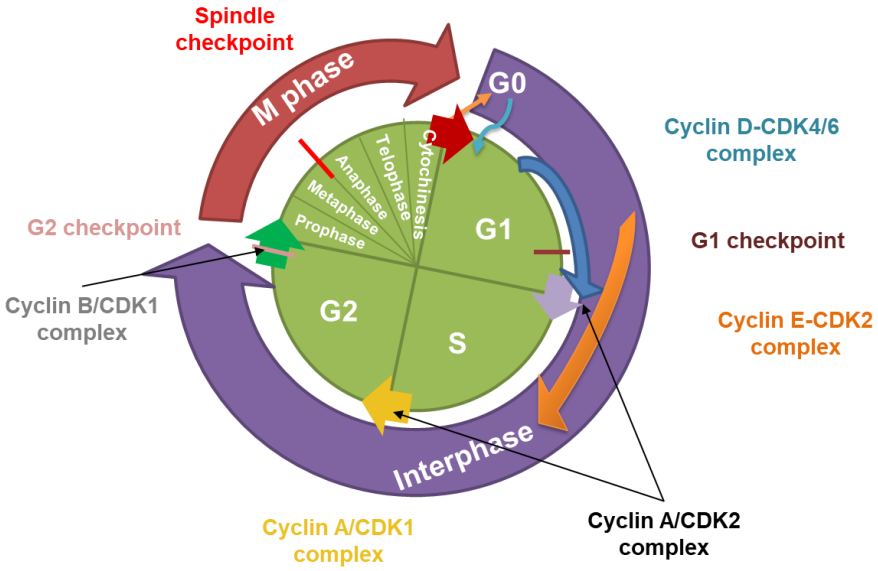


Figure 1.7: The mammalian cell cycle.

A schematic diagram of the cell cycle phases with the cdks and cyclins that control the transition between the phases. The diagram also shows positions of the checkpoints at which the cell cycle might be arrested enabling repair and maintenance of a cell before it progresses down the cycle. Image modified the University of Tokyo *http://csls-text.c.u-tokyo.ac.jp/active/09_06.html* and (Malumbres and Barbacid, 2009).

If any cell receives a signal to proliferate, which usually comes from the microenvironment, it leaves the G0 phase and transits to G1 phase. In the G1 phase, a cell prepares for DNA replication in the next phase. Here, CDK4 and CDK6 form complexes with cyclin Ds (1, 2 and 3) and get activated by the CDK activating enzyme (CAK) cdk7/cyclin H/MAT1. Progression from G1 to S phase is controlled by the first restriction point of the cell cycle that is dependent on the retinoblastoma susceptibility protein (Rb), a key negative regulator of the G1-S transition. Rb is a transcriptional co-repressor that binds transcription factors of the E2F family, and represses E2F target genes. For the G1-S transition to be accomplished, CDK4/6-Cyclin D and CDK2-Cyclin E sequentially phosphorylate the Rb protein, which eventually result in the E2F-TF release for the induction of transcription of the vital genes for entering the DNA replication (S) phase. After the transition, the activities of the complexes CDK2-Cyclin A, CDK1-Cyclin A and CDK1-Cyclin B keep the Rb protein hyperphosphorylated throughout the cell cycle, whereas, the cyclins D1 and E are degraded through ubiquitin-proteasome pathway early in S phase (Bachs et al., 2018; Casimiro et al., 2014; Malumbres and Barbacid, 2009).

1.4.4.2 Cell cycle inhibitors (CDK-Is) and the G1 checkpoint

In the absence of proliferative and growth signals, the progression of the cell cycle is inhibited by the CDK-Is. The reason for the existence of these mechanisms is to make sure that the cell cycle progression is not permitted unless all the requirements for progression to the next phase are fulfilled. Upon the exposure of the cells to growth signals, cell cycle effectors repress CDK-Is transcription, or induce their ubiquitination-dependent degradation, which results in cell cycle progression.

1.4.4.2.1 The G1 checkpoint (Restriction point)

At G1 phase, a cell is thought to set up for DNA replication. This is achieved by accumulation of the proteins required for DNA replication, as well as the assembly of replication machinery at the origins of DNA replication. Additionally, a cell ensures its responsiveness to growth signals and nutrients supporting cell division. Of the whole cell cycle, this is the only stage whereby a cell is sensitive to EC environment-released growth signals; and growth factor dependency is believed to diminish upon a cell progresses through the G1 or restriction point (Barbash and Alan Diehl, 2008; Yao, 2014).

The G1 check- or restriction point is governed by members of the Rb family of tumour suppressor proteins (aka pocket proteins), which are Rb, p107 and p130. They control proliferation of several cell types and they do so by interacting with the E2F family of transcription factors. The E2F family consists of transcriptional activators (E2F1-3a) and repressors (E2F3b-8) and are implicated in the transcriptional control of genes implicated in cell cycle progression and DNA replication. At quiescence, the Rb proteins associate with the E2F activators and prevent their DNA binding and transcriptional activity. Additionally, the Rb proteins interact with the E2F repressors forming complexes that enhances their repressive functions by recruiting the several chromatin modifying proteins (Cobrinik, 2005; Yao, 2014).

For a cell to transit from G1 phase to the S-phase, upon exposure to growth signals, the Rb proteins must be subjected to a series of phosphorylation events. These phosphorylation events are sequentially performed by the cyclin D-CDK4/6 and cyclin E-CDK2 complexes, as previously mentioned (section

1.4.4.1). These phosphorylation events cause modifications in the Rb proteins conformation that culminates in their dissociation from the E2F activators, thus enabling them to activate transcription (Barbash and Alan Diehl, 2008; Coller, 2007)(Figure 1.8). At mid G1 phase, the cyclin D-CDK4/6 complexes are thought to interact with and sequester the CDK-Is p21Cip1 and p27Kip1, thus, blocking their interactions with CDK2. This enables CDK2 to interact with cyclin E, forming the CDK2-cyclin E complexes that phosphorylate Rb. This stage (i.e. after E2F activators release) is thought to demarcate the G1-S transition and the end of cell cycle progression sensitivity to the growth signals (Barbash and Alan Diehl, 2008).

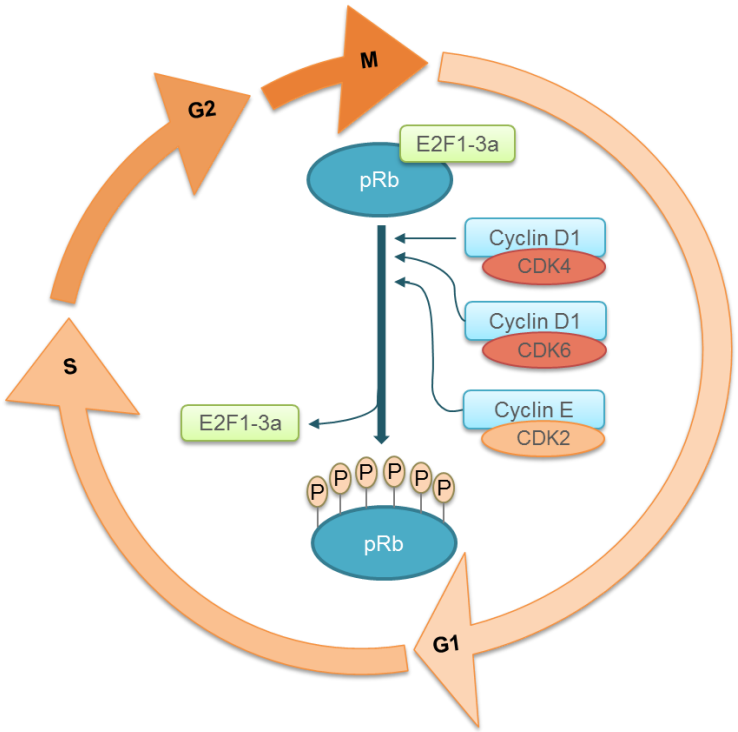


Figure 1.8: A schematic representation of Rb regulation at the G1-S phase transition.

The Rb (pRb) proteins are initially phosphorylated by the cyclin D-CDK4/6 complexes and then by the cyclin E/CDK2 complex resulting in a change in its conformation. This causes its dissociation, the release of the E2F activators (E2F1-3a) and the subsequent activation of target genes including cyclin E gene. The scheme is adapted from (Coller, 2007)

1.4.4.2.2 The CDK-Is

The CDK-Is function is to inhibit the progression of the cell cycle beyond G1 through blocking the activity of cyclin/CDKs complexes. Based on their structure and function, the CDK-Is are divided into two groups, the INK4 family and the Kip/Cip family (Malumbres and Barbacid, 2001; Musgrove et al., 2004). The INK4 family consists of $p16^{INK4A}$, $p15^{INK4B}$, $p18^{INK4C}$ and $p19^{INK4D}$. Those are known to bind CDK4/6 with high specificity, which results in the dissociation of cyclin D and the inhibition of the kinase activity. Unlike the INK4 family, the Kip/Cip family members, which are p27Kip1, p21Cip1 and p57Kip2, interact with both, cyclins and CDKs. They are also able to bind both cyclin D-CDK4/6 as well as the cyclin E/A-CDK2 complexes, although with different affinities. For example, p27Kip1 concentration that is required for cyclin E/A-CDK2 inhibition is lower than the cyclin D-CDK4/6 inhibiting concentration. Kip/Cip family members are considered as intrinsically unstructured due to the absence of a stable tertiary structure upon isolation. The cyclin/CDK binding domain is believed to be conserved among the three proteins and is located in the N-terminus. However, the C-termini of the proteins contain motifs for post translational modification and protein-protein interactions, and are less conserved (Musgrove et al., 2004; Sharma and Pledger, 2016).

The protein p27Kip1is encoded by the gene *CDKN1B* and is a 198 amino acids (aa) long (Bencivenga et al., 2017; Chang et al., 2004). In addition to the cyclin E-CDK2 binding, p27Kip1 is capable of interacting with cyclin D-CDK4/6, cyclin A-CDK2/1 complexes and cyclin B-CDK1. Like other members of the Kip/Cip family, its N-terminal part contains various well-structured domains that are collectively known as the kinase inhibitory domain (KID), which occupies the 25-89 aa area (Figure 1.9 A). The KID consists of a subdomain for cyclin-binding (D1), the subdomain for CDK-binding (D2) and a subdomain that links D1 and D2 (LD). KID also contains a signal for nuclear export (NES) between aa 32 and 46. Adjacent to the KID lies a region that is proline rich (aa 90-96), which is implicated in the interaction with Grb2 adaptor protein. As previously mentioned, the C-terminal domain is much less structured and is believed to interact with various proteins. It also harbours a signal for nuclear localisation (NLS) localised within the 152-168 aa area (Bachs et al., 2018; Sharma and Pledger, 2016).

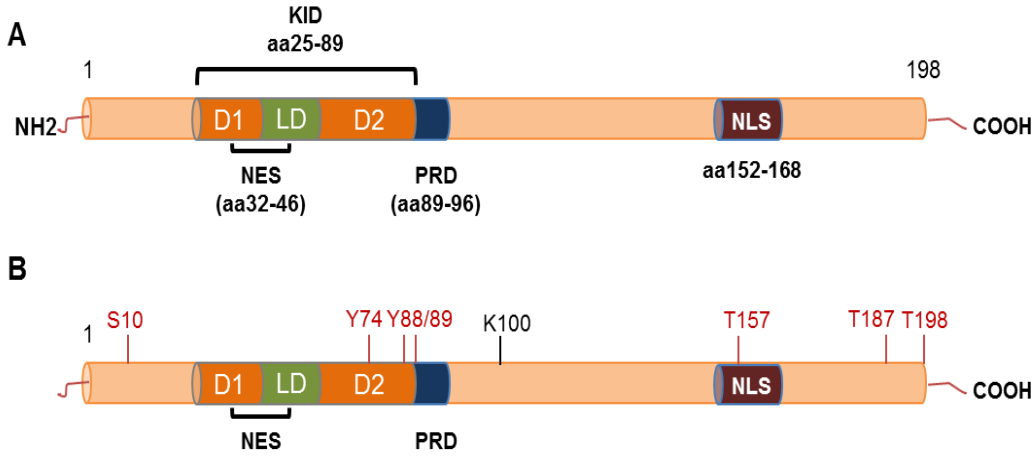


Figure 1.9: The schematic representation of the p27Kip1 protein.

A shows the main functional domains of the protein and their location within the protein sequence. **B** highlight the major sights for the post-translational modification of the protein. The scheme is adapted from (Bachs et al., 2018).

Mechanisms of CDK inhibition by p27Kip1 are well understood. A part of D2 of p27Kip1 interacts with the CDKs catalytic pocket, hence, competing with ATP binding and inhibiting the CDK activation. During G1 phase, the tyrosine residues (Y74 and Y88) of D2 are phosphorylated by several tyrosine kinases, such as Src kinases, which modifies the D2 conformation and release its localisation allowing the CDK to be partially active (Figure 1.9 B). The partially active CDK then further phosphorylates p27 at the threonine residue 187 (T187), which induces its dissociation and Skp2 dependent ubiquitination and degradation leading to the full activation of the CDK (Bachs et al., 2018). Another mechanism through which p27Kip1 degradation is induced involves acetyl transferase p300/CBP associating factor (PCAF)-mediated acetylation of lysine 100 (K100) resulting in nuclear degradation of the protein. Phosphorylation of Serine 10 residue (S10) by various kinases, such as cdk5, is thought to induce the protein's translocation to the cytoplasm. Subsequent phosphorylation at residues T157 and T198 retains p27Kip1 in the cytoplasm where it can be either stabilised or degraded in an E3 ubiquitin ligase dependent manner (Belletti and Baldassarre, 2012; Sharma and Pledger, 2016).

p27Kip1 is a multifunctional protein, its functions range from the canonical cell cycle regulation to the non-canonical functions in cell migration and transcription. The canonical cell cycle regulation involves inducing and maintaining the cell quiescence. p27Kip1 is a tumour suppressor, even though the mutations are very rare. However, abnormal level or localisation of p27Kip1 was associated with cancer and neurodegeneration (Barbash and Alan Diehl, 2008; Chu et al., 2008; Garrett-Engele et al., 2007). For instance, p27Kip1 expression loss has been shown to contribute to genetic instability and radioresistance in luminal breast cancer cells (Berton et al., 2017). Additionally, it has been reported that cytoplasmic p27Kip1 is implicated in the reorganisation of the cytoskeleton (Besson et al., 2004b).

1.4.5 EMT-TFs and genetic and chromosomal instability

The accumulation of mutations and genetic instability is considered as key empowering hallmark of cancer progression and a main source of cancer evolution, as previously stated (section 1.1). Several studies have linked EMT induction and EMT-TFs to the accumulation of genetic instability. Comaills et al. (2016), for example, has shown that TGF-β and SNAIL induced EMT resulted in anomalous mitosis leading to aneuploidy. They have found that EMT in the normal human breast epithelia MCF-10A, the human breast cancer cell line SKBR3 and the mouse mammary epithelial cell line comma-D β-geo (CD β-geo) resulted in the elevation of the number of binucleated (BN) cells. These BN cells were found to suffer clustering of their centrosomes and bipolar division, which then led to missegregation of the chromosomes, cytokinesis failure and micronuclei (MN) generation. This was found to considerably damage the misseggregated chromosomes. Additionally, EMT-induction in these cells resulted in nuclear blebbing and reduced the circularity of the nucleus. Mass spectrometric analysis of the proteins from these cells has revealed a severe impairment and weakening of the nuclear envelope (NE) due to a reduction in NE proteins expression, such as laminB1. This was then suggested to reduce the ability of the nucleus bear the mechanical forces imposed on a cell as it migrates. This can eventually result in the rupture of the nucleus leading to the release of the DNA and other nuclear components to the cytoplasm and the augmentation of the damage. As a consequence, this could result in severe

effects on the integrity of the genome and the accumulation of mutations and eventually, fragmentation and chromothripsis (Lee et al., 2016). EMT induction in these cells was found to be reversible, however, the deleterious genetic instability was heritable, suggesting a potential contribution to tumorigenesis. In addition, Bakhoum et al. (2018) demonstrated that EMT is induced in genetically unstable cancer cells via a mechanism involving spillage of genomic DNA into the cytosol, activation of the DNA-sensing interferon pathway, and noncanonical NF-κB signalling. This work suggests that the defects in chromosomal segregation may cause metastatic spread via EMT pathways.

On the other hand, EMT induction and EMT-TFs has been implicated to serve a protective role on genomic integrity. A study using radioresistant breast cancer cells has shown that the EMT-TF ZEB1 is phosphorylated and stabilised by the action of the Ataxia Telangiectasia Mutated (ATM) kinase following ionizing radiation and DNA damage checkpoint activation (Zhang et al., 2014b). This stabilisation then enables ZEB1 to interact with and facilitate the functioning of the deubiquitylation enzyme USP7. USP7 then deubiqutylates and stabilises the effector kinase CHK1. This in effect enhances the ability of these radioresistant cells to repair their double strand DNA breaks by homologous recombination (HR) and enhance cell survival (Zhang et al., 2014b). In parallel, activated ZEB1 also interferes with its own inhibition by downregulating the expression of its negative regulator miR205. This in turn results in a further elevation of ZEB1 expression as well as an increase in the expression of Ubc13. Ubc13 is an ubiquitin-conjugating enzyme that is critical for HR initiation by recruiting RPA, BRCA1 and RAD51 to the DSB site (Zhao et al., 2007). ZEB1 and Ubc13 upregulation further increases DNA repair by HR, hence, reinforce radioresistance (Zhang et al., 2014a). Another study has suggested that cyclin E or Ras oncogenic transduction in non-transformed mammary stem cells resulted in the initiation of a ZEB1- methionine sulfoxide reductase (MSRB3) axis, which enables them to resist oncogene-induced DNA damage (Morel et al., 2017) . *MSRB3* gene is a direct transcriptional target of ZEB1, and its elevated expression reduces the level of reactive oxygen species (ROS) in these cells. Subsequently, in response to oncogenic insults, mammary stem cells ($CD44^{high}CD24^{low}$) developed much less DNA damage and chromosomal

instability as compared to committed or differentiated cells ($CD44^{high}CD24^{high}$ or $CD44^{low}CD24^{high}$). So, in this study, it has been suggested that ZEB1-positive mammary stem cells are more susceptible to oncogenic transformation, because ZEB1/MSRB3 pathway maintains genome integrity in those cells and protects them from ROS-induced apoptosis.

EMT also correlates with chromosomal instability. For instance, Twist overexpression was observed in tumour samples isolated from breast cancer patients and this correlated with invasiveness and chromosomal instability (Mironchik et al., 2005). Additionally, TGF-β induced EMT in proliferating MCF-10 and Cdβgeo cells resulted in chromosomal abnormalities and chromosome number variations that resulted from chromosomal nondisjunction and disrupted mitotic spindles. Furthermore, EMT and TGF-β signature positive CTCs that are isolated from the blood of women with metastatic breast cancer and from men with metastatic prostate cancer showed a correlation with aneuploidy (Comaills et al., 2016)

1.5 DNA damage response

DNA Damage response (DDR) refers to the mechanisms that are implicated in eliminating damages that could occur in the genome. This involves the interaction and cooperation between components of multiple complex pathways implicated in cell cycle checkpoints, DNA damage repair, chromatin remodelling and apoptosis. This is to ensure the maintenance of the genome fidelity. It is believed that the human genome is continuously exposed to various forms of deleterious insults that can, if not repaired, interfere with DNA replication and transcription and can lead to accumulation of mutations, if misrepaired. DDR importance is apparent because different diseases such as Fanconi anaemia and cancer are associated with DDR deficiency or malfunction, (Ali et al., 2017; Friedberg, 2003b; Jackson and Bartek, 2009; Lei and Lee, 2005).

1.5.1 Types of DNA damage

DNA insults are either endogenous (a result of normal metabolic processes within the cells), or exogenous (induced by environmental factors). An example of endogenous insults is the spontaneous acid-catalysed hydrolysis of the N-glycosidic bond between the deoxyribose backbone and the DNA base, resulting

in the generation of abasic or apurinic/apyrimidinic (AP) sites. Thousands of AP sites are generated per cell every day. In addition, endogenous insults might arise from mismatches, deletions and insertions that are generated by DNA polymerases during DNA replication (Jackson and Bartek, 2009; Marnett and Plastaras, 2001).

Examples of exogenous DNA insults include the ultraviolet (UV) light and ionizing radiation. UV light is considered to be the most prominent form of environmental agents that can damage the DNA. It is expected that for every hour of exposure to strong sunlight, A and B forms of UV light can generate approximately 100,000 DNA lesions for each exposed cell. UV light results mostly in the generation of pyrimidine dimers by forming aberrant covalent bonds between adjacent pyrimidines. Ionizing radiation (IR) is thought to generate various types of DNA lesions, the most harmful of which are DSBs. IR can either be generated following natural radioactive compounds decay, such as the production of radon gas from uranium decay, or can be man-made, such as radiotherapy and X-rays. In addition to inducing direct damage to the DNA, UV and IR can indirectly cause DNA damage, through the generation of reactive oxygen species (ROS) (Basu, 2018; Hoeijmakers, 2001; Jackson and Bartek, 2009). ROS are known for generating over 100 forms of DNA adducts, including bases alterations, oxidation of the deoxyribose backbone, as well as the generation of single and double strand breaks (SSB and DSB).

1.5.2 Mechanisms of DNA repair

Each type of DNA lesions is thought to activate a particular mechanism of the DDR, which in turn is responsible for the elimination and repair of that insult. In mammals, at least 5 main pathways of DNA damage repair exist, which are often co-regulated and partially overlap. These pathways are Base excision repair (BER), Nucleotide excision repair (NER), Mismatch repair (MMR), Homologous recombination (HR) and non-homologous end joining (NHEJ).

1.5.2.1 Repair of single strand damage

1.5.2.1.1 Base Excision repair (BER)

BER is the mechanism that repairs by merely removing the affected base by cleaving the N-glycosidic bond that connects it to the deoxyribose sugar

backbone. This process is mainly implicated in repairing DNA lesions that arise due to endogenous damaging sources like oxygen radicals (i.e. damages that do not distort the DNA double helix). Once a base is modified, it is recognised by a group of enzymes, known as the DNA glycosylases, which would then initiate BER by removing the damaged base and forming an abasic site in the DNA strand. Following this, apurinic/apyrimidinic endonucleases associate with and nick the DNA by cleaving the phosphodiester bond adjacent to the abasic site resulting in the excision of the damaged nucleotide. After that, missing nucleotide(s) are incorporated into the damaged site by DNA polymerases and finally, DNA backbone gaps are ligated by the DNA Ligases (D'Andrea, 2015; Hoeijmakers, 2001; Wallace, 2014).

1.5.2.1.2 Nucleotide excision repair (NER)

The process of NER involves the removal of the whole nucleotide (including the modified base) from the DNA helix. NER is thought to be implicated in the repair of alterations that would distort the double helix of the DNA, which are usually exogenously derived, such as thymine dimers induced by UV. Generally, NER is thought to repair lesions that would inhibit base-pairing and cause transcription and replication blockage. Two subpathways of NER are thought to exist, one that safeguard the whole genome and is known as global genome NER (GG-NER). The other subpathway is merely implicated in repairing lesions that would inhibit the elongation phase of transcription, and is known as transcription coupled repair (TCR)(D'Andrea, 2015).

Briefly, upon the formation of a thymine dimer, the damage is initially recognised by a sensor complex and a damage specific DNA binding protein (DDB). DDB then recruits the excision repair complex, whose components help to unwind the DNA double helix and cleave DNA on the 3' side as well as the 5' side of the damage. This double incision results in the release of the DNA-sequence containing the damage, typically ~24 nucleotides long. Next, the excised sequence is resynthesized with the help of Proliferating Cell Nuclear Antigen (PCNA), and the DNA backbone is finally religated (D'Andrea, 2015; Friedberg, 2001).

Mutations in the genes that encode for components of the nucleotide excision repair complex underlie several congenital diseases. An example of such diseases is xeroderma pigmentosum, which is a skin disorder characterised by an extreme sensitivity to the sun light. People with this disorder are highly predisposed to the development of skin cancer (D'Andrea, 2015; Hoeijmakers, 2001).

1.5.2.1.3 Mismatch repair (MMR)

MMR repair process is primarily involved in the fast elimination of mispaired nucleotides that are generated by DNA polymerases during replication. MMR is also responsible for insertions or deletions of slipped nucleotides during recombination or the replication of highly repetitive sequences. In addition, MMR is thought to be implicated in DNA adducts recognition and repair. MMR consists of 4 steps; detection of the mismatch, additional MMR factors recruitment, excision of the mismatch and finally restoring the excised portion of the sequence by copying the undamaged complementary DNA strand (D'Andrea, 2015; Modrich, 2006).

In comparison to normal cells, MMR deficient cells display an elevated spontaneous mutation rate that increases the risk of malignant transformation. For instance, germ line mutations of MMR genes have been recognised as predisposing factors for the hereditary form of colon cancer (hereditary nonpolyposis colon cancer, HNPCC), as well as other sporadic cancers (Aaltonen et al., 1993; Friedberg, 2003a; Vasen et al., 2013).

1.5.2.2 Repair of double strand breaks (DSB)

DNA DSBs are the most deleterious of the DNA lesions, which if left unrepaired, are implicated in the direct induction of cell death or genetic instability. DSBs are thought to be generated as a result of IR, ROS, genotoxic chemicals (such as chemotherapeutics) and as a result of SSBs replication. Upon the formation of a DSB, a sophisticated network of pathways activates DNA damage checkpoints, cell cycle arrest and DSBs repair machinery (D'Andrea, 2015; Hakem, 2008; Hoeijmakers, 2001).

DSBs can also be deliberately generated as a result of normal physiological functions and are known as programed DSBs. An example of such functions is

generation of DSBs during reassembly of gene segments in the V (variable), D (diversity), and J (joining) (V(D)J) recombination for the generation of diverse antigen receptors in T- and B-cells (Roth, 2014). Like accidental DSBs, programed DSBs can also lead to the accumulation of mutations, chromosomal aberrations and ultimately cell death if V (D) J recombination fails.

There are two main pathways for the repair of DSBs, the process of homologous recombination (HR) and the process of non-homologous end joining (NHEJ). The choice of the repair pathway is regulated throughout the cell cycle (Hakem, 2008).

1.5.2.2.1 Homologous recombination (HR)

HR repair of DSBs essentially involves the alignment of the damaged sequence with the intact homologous sequence. This process is mainly active between late S-phase and M-phase of the cell cycle (i.e. after the DNA is replicated and used as a template for the damaged DNA). HR repair is considered to be a more accurate and slower form of DSB repairing mechanisms as compared with NHEJ (see below). It is also of a particular importance in repair of inter-strand crosslinks, collapsed replication forks and incomplete telomeres. HR is also involved in the genetic material exchange between the homologous chromosomes and the proper segregation of the homologous chromosomes pairs during meiosis (Jasin and Rothstein, 2013).

HR comprises 3 fundamental steps, which are the invasion of the strand, migration of the branch and the formation and resolution of a Holiday junction (HJ). Firstly, the DNA ends at the DSB are resected by the MRN complex, which is composed of the Mre11, Rad 50 and Nbs1 as well as few accessory proteins, such as the tumour suppressor BRCA1, CtBP-interacting protein (CtIP), the Bloom helicase, the exonuclease 1 (Exo1) and the helicase/nuclease Dna2 (Figure 1.10). This induces the removal of the 5’ end terminal nucleotides at both sides of the DSB and the generation of single-strand DNA (ssDNA) overhangs at the 3’ end, which are the substrates for the HR machinery. After that, the replication protein A (RPA) trimer covers the ssDNA overhangs to stabilise them, blocks their nucleolytic cleavage and prevents the formation of secondary structures. RPA also triggers the heptamer Rad51 recombinase to displace it

and, in a cooperation with other accessory proteins such as BRCA2, forms nucleoprotein filaments. Rad51 then, with the support of branch-migration activity of Rad54, is thought to catalyse the search for homology on the sister chromatid, strands-pairing and exchange and ultimately the formation of the displacement loop (D-loop). After this step, while using the donor strand as a template, DNA is synthesised by polymerases concomitant with Rad51 removal. From here, two modes of HR can take place, in the first one that is known as Synthesis-Dependent Strand Annealing (SDSA), merely one of the 3'-ssDNA is involved in the process resulting in a single HJ formation. However, in the second mode of HR that is known as Double-Strand Break Repair (DSBR), both 3'ssDNA are involved in the process and hence, two HJ are formed. Ultimately, the HJ are resolved by resolvases (e.g. GEN1) culminating in the restoration of the original DNA sequence around the DSB and this can also involve crossover (Dueva and Iliakis, 2013; Jasin and Rothstein, 2013).

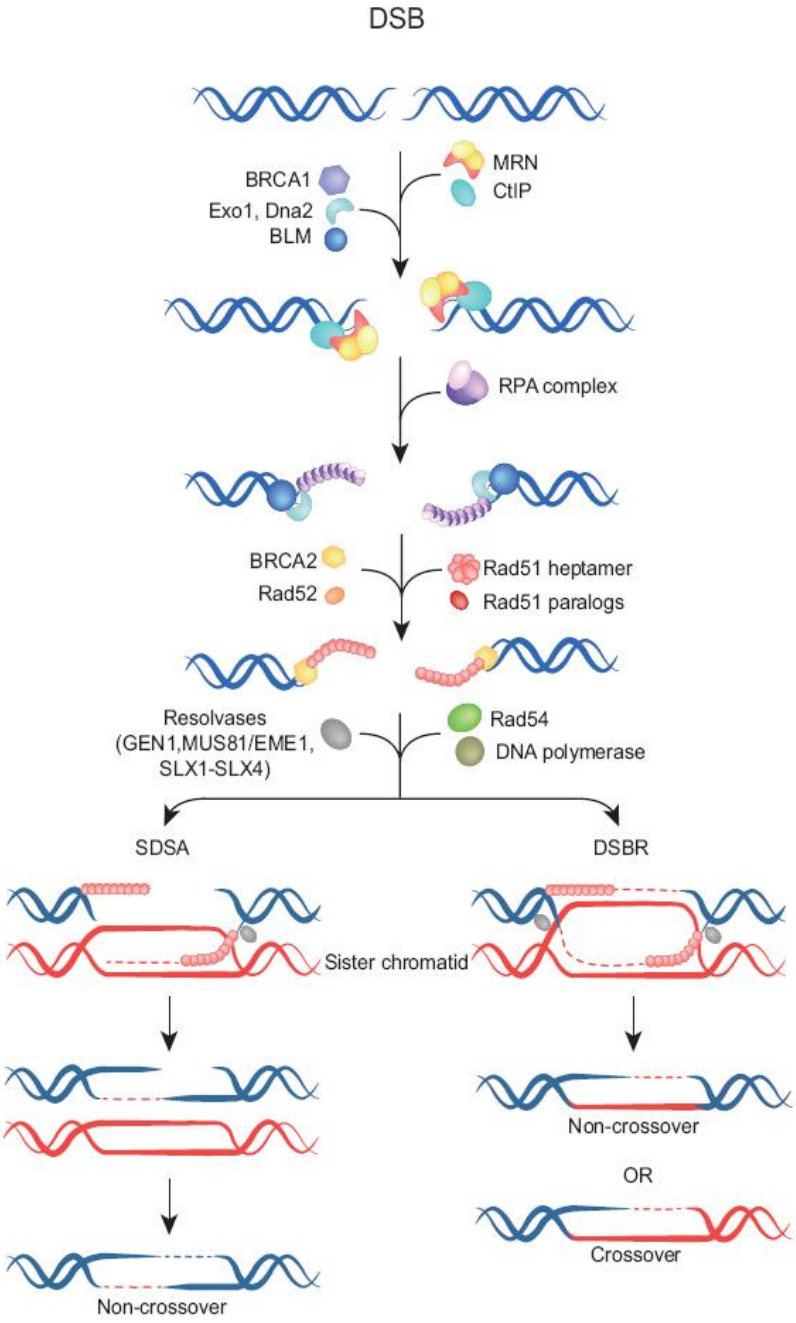


Figure 1.1022: A schematic representation of the HR repair of DSBs.

The scheme illustrates the steps of the HRR process and the enzymes involved. It also shows the two modes of HR: the synthesis-dependent strand annealing (SDSA) and the double-strand break repair (DSBR). The difference between the two modes is that the SDSA involves the use of only one 3'-single-strand DNA (ssDNA) overhang and only a single Holiday Junction (HJ) is formed. In comparison, the DSBR makes use of both 3'-ssDNA overhangs where two HJ are formed and this is the one that could result in crossovers. The scheme is taken from (Dueva and Iliakis, 2013)

Although HR is thought to be implicated in the repair of only 10% of DSBs in mammalian non-meiotic cells, germline mutations that affect this pathway result in deleterious outcomes. These mutations have been shown to cause the development of various diseases and increase risk for cancer.

For example, mutations linked to the MRN complex (e.g. *MRE11* loss of function mutation) have been shown to cause the rare ataxia telangiectasia (AT)-like disorder (ATLD). Patients with ATLD suffer from immunodeficiency and progressive neurodegeneration. At the cellular level, loss of MRE11 function causes abnormality in the repair of DSB and downstream cellular responses. ALTD is also linked to an increased IR sensitivity, increased level of unrepaired DNA damage, and genomic instability (Taylor et al., 2004).

Similarly, hypomorphic mutations in *NBN*, the gene that codes for the MRN complex component NBS1, result in the development of the rare autosomal recessive disease NBS. The symptoms are: microcephaly, immunodeficiency and other growth abnormalities. At the cellular level, *NBN*-deficient cells exhibit hypersensitivity to DNA insults, accumulation of endogenous- and exogenous-induced DNA damage, chromosomal instability and cell-cycle checkpoints defective activation. Additionally, NBS patients are predisposed to cancer, especially leukaemia (Chrzanowska et al., 2012).

Breast cancer early onset (BRCA) proteins 1 and 2 (BRCA1 and BRCA2) play a key role in HR DNA repair. Mutations in these genes are linked to chromosomal instability and cancer predisposition. Germline mutations in *BRCA1* and *BRCA2* are believed to predispose to multiple forms of cancers, such as the hereditary form of females' breast and ovarian cancers and to prostate cancer (Campeau et al., 2008; Pal et al., 2005). BRCA1 and BRCA2 have also been implicated to the development of the prevalent familial form of bone marrow failure, Fanconi anaemia (FA). Patients with FA suffer progressive failure of the bone marrow that usually starts with pancytopenia, which is the decline in haematopoietic cells numbers. These patients also suffer from a higher sensitivity to DNA damage and the accumulation of genomic instability that ultimately predisposes them to cancer development (Ceccaldi et al., 2016).

Upon the DNA damage checkpoint activation, ATM, ATR and Chk2 phosphorylate BRCA1. Phosphorylated BRCA1 then interacts with the MRN complex and ɣ-H2A.X at the DSB where they initiates its resection (Maréchal and Zou, 2013). In addition, BRCA1 has important functions that are HR independent. It is thought to be involved in oestrogen receptor signalling, ubiquitination, chromatin remodelling and transcription (Deng, 2006; Liu et al., 2008a; Narod and Foulkes, 2004). BRCA2 interacts with Rad51 resulting in its recruitment and loading for the formation of the nucleoprotein filaments at the 3'-ssDNA overhang. Additionally, BRCA2 has recently been implicated in stabilising stalled replication forks and the subsequent blockage of replication progression (Lomonosov et al., 2003).

1.5.2.2.2 Non-homologous end joining (NHEJ)

The other key pathway for the repair of DSBs is the NHEJ, which merely joins the DNA ends at the DSBs. This process is more active in non-proliferating cells and mainly at G1 phase of the cell cycle (i.e. before the sister chromatids are synthesised). This process is considered more prone to errors because, firstly, unlike HR, it lacks the intrinsic potential for restoring the DNA sequence near the DSB. Secondly, NHEJ lacks the intrinsic potential for restoring the DNA molecule to its original state, where it technically can join any two DNA molecules in the vicinity. These in essence could result in the generation of various modifications in the DNA sequence that could consequently lead to deleterious mutations. However, due to the extremely high operational speed of the process, with approximately 10-30min half time, NHEJ is regarded as the genomic integrity guardian and is thought to be associated with a lower probability of generating sequence alterations. NHEJ is also thought to play a vital role in the V(D)J recombination of antigen receptors genes and class switch recombination (CSR) in T-cells, hence, contributing to their diversity (Davis and Chen, 2013; Dueva and Iliakis, 2013; Roth, 2014).

In general, classical NHEJ (C-NHEJ) process at the DSB is composed of 4 steps, which are the recognition of the broken DNA end and NHEJ complex formation and stabilisation. This is then followed by broken end bridging and stabilisation, processing and finally ligation and NHEJ complex dissociation. So, upon the formation of a DSB, C-NHEJ is activated, where the blunt end binding proteins

Ku70/Ku80 heterodimer recognises the broken DNA ends and create a ring that circles the broken end (Figure 1.11). This circulation inhibits any non-specific resection of the broken end and recruits the catalytic subunit of the DNA-protein kinase (DNA-PKcs), which in turn phosphorylates and recruits the exonuclease Artemis to process and resect the broken end. DNA-PKcs-Ku complex also recruits and activates the DNA ligase IV (LIG4) and other accessory factors like repair cross-complementing protein 4 (XRCC4) complex and XRCC4-like factor (XLF) that ultimately ligate the broken ends. The NHEJ repair complex finally dissociates from the ligated and repaired DNA molecule (Davis and Chen, 2013; Dueva and Iliakis, 2013).

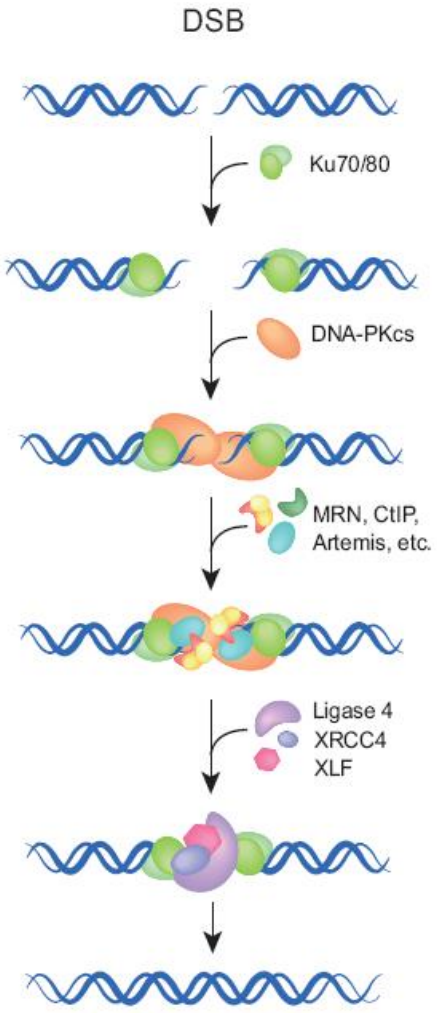


Figure 1.1131: A schematic representation of the NHEJ repair of DSBs.

The scheme illustrates the steps of the NHEJ process and the factors involved. Upon the generation of a DSB, the Ku heterodimer recognises and encircles the DNA ends at the DSB. This in turn recruits the catalytic subunit of the DNA-PK (DNA-PKcs), where it complexes with Ku heterodimer. Ku-DNA-PKcs complex then recruits and activates other components of the NHEJ complex, including Artemis, DNA Ligase IV, XRCC4 and XLF. These factors then process and finally ligates the broken ends. The scheme is taken from (Dueva and Iliakis, 2013)

Mutations in the components of the NHEJ repair pathways have been shown to lead to syndromes with severe immunodeficiency and cancer predisposition. Mutations of the nuclease Artemis, for instance, are associated with the development of the rare human disorder of severe combined immunodeficiency with ionizing radiation sensitivity (RS-SCID). Patients with RS-SCID suffer defective NHEJ (defects in VDJ recombination) that results in a blockage of T- and B-cell lymphocytes differentiation (O'Driscoll and Jeggo, 2006).

1.5.3 The DNA Ligases

DNA ligases are ATP-dependent nucleotidyltransferases that catalyse the formation of the phosphodiester bonds in the DNA strands by ligating adjacent 3'-hydroxyl and 5'-phosphoryl termini. In this reaction, initially, an adenylate group (adenosine 5'-monophosphate or AMP) is serially moved from ATP to a lysine residue in the active site of the enzyme. The AMP is then transferred from the enzyme's active site to the 5' phosphate end of the DNA. This adenylation of the 5' phosphate results in its activation and prepares it for the 3' hydroxyl nucleophilic attack, which culminates in AMP displacement and the covalent joining of the two DNA ends.

Eukaryotes have three families of DNA Ligases (I, III and IV). The families of DNA ligase I and IV are thought to be found in all eukaryotes whereas the DNA ligase III family is restricted to vertebrates only (Arakawa and Iliakis, 2015). Members of the three families share a conserved catalytic core region that consists of a DNA-binding domain (DBD), an adenylation or a nucleotidyltransferases domain (AdD or NTase) and an oligonucleotide/oligosaccharide binding (OB)-fold domain (OBD) (Ellenberger and Tomkinson, 2008; Singh et al., 2014). The catalytic core region is thought to circle the DNA nick where each of the domains is in association with the DNA double helix (Pascal et al., 2004). The flanking non-catalytic regions instruct the DNA ligases localisation and function within the cell (Howes and Tomkinson, 2012). The DNA ligase I is a 120kDa enzyme that is composed of 919aa and is encoded by the *LIG1* gene. On the N-terminus, the enzyme harbours a replication factory-targeting sequence (RFTS) between residues 2 and 9 of the

protein (Figure 1.12). RFTS also functions as a PCNA interacting peptide (PIP) box that allows DNA ligase to interact with PCNA. 4 serine residues (S51, S66, S76 and S91) are posttranslationally modified. Phosphorylation of these residues by casein kinase II and CDKs during the cell cycle progression keep LIG1 in a hyper-phosphorylated form that is essential for its function in the replication. The protein then contains a nuclear localisation sequence (NLS) between residues 111 and 179. This is followed by a DBD between residues 262 and 534 and a core catalytic region between residues 535 and 919, which harbours an AdD (aa 535-748) and an OBD (aa 749-919). The AdD contains the active site (Lys, K568) which forms the covalent bond with AMP (Ellenberger and Tomkinson, 2008; Howes and Tomkinson, 2012; Singh et al., 2014).

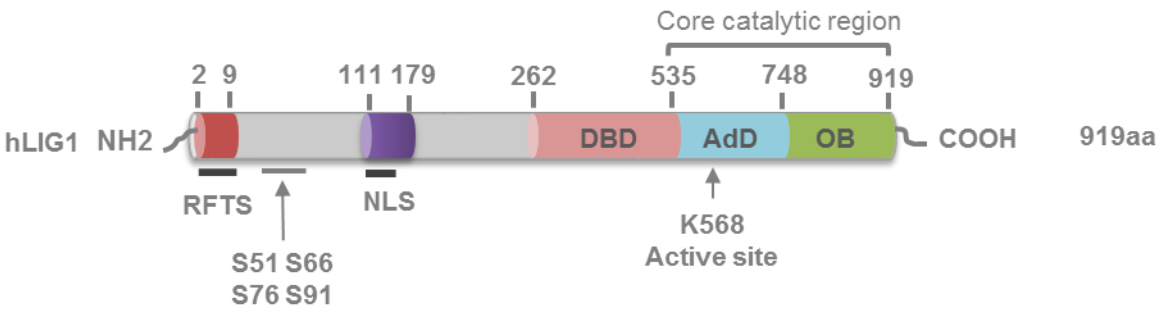


Figure 1.12: A schematic representation of LIG1.

The scheme shows the main functional domains of the protein and their location within the protein sequence. It also highlights the major sights for the post-translational modification of the protein as well as the active site within the adenylation domain (AdD). DNA-binding domain (DBD), oligonucleotide/oligosaccharide binding domain (OBD), replication factory-targeting sequence (RFTS), nuclear localisation sequence (NLS).

The DNA Ligases are known for their indispensable role during the processes of DNA repair, recombination and replication. LIG1, for instance, is thought to serve an important function in the ligation of the Okazaki fragments of the lagging strand during DNA replication. It has also been shown to be very important for the ligation step in long patch BER, NER, HR and microhomology end joining (MHEJ)(Liang et al., 2008). On the other hand, LIG3 is the only ligase implicated in replication and repair of mitochondrial DNA, and in the ligation step of long patch BER. LIG4 is the only ligase implicated in the repair of the DSB by the C-

NHEJ as well as in V (D)J recombination and CSR (Arakawa and Iliakis, 2015; Ellenberger and Tomkinson, 2008).

Recently, LIG3 has been suggested to be capable of compensating for some of LIG1 functions such as ligation of Okazaki fragments, NER, HR and MHEJ. However, LIG3 seems to be unable to compensate for LIG1 function in the maintenance of telomeric integrity (Arakawa et al., 2012; Arakawa and Iliakis, 2015; Le Chalony et al., 2012)

1.5.4 DDR and EMT

Previous studies in our lab have shown that ZEB2 attenuates the UVC-induced DNA damage response. Alkaline comet assays revealed persistence of single strand breaks after the exposure of A431-ZEB2 cells to low dose of UVC (Dr Hussein Abbas, Professor G Don Jones lab, University of Leicester). They have found that repairing the gaps in the DNA backbone by DNA polymerase and the DNA ligase I were strongly reduced, but the efficacy of the excision of modified bases was not affected. Quantitative-PCR microarray demonstrated that ectopic expression of ZEB2 in A431 squamous cell carcinoma cells affected the transcription of approximately 50% of DNA damage response genes tested (42/94) (Dr Gina Tse, Dr Eugene Tulchinsky lab, University of Leicester, unpublished data). One of the candidate genes was the DNA Ligase I, which probably could explain the reduction in DNA backbone gap repairing shown previously. After that, downregulation of DNA Ligase I expression by ZEB2 was further confirmed through western blotting (Dr Gina Tse, Dr Eugene Tulchinsky lab, University of Leicester, unpublished data). It was found that DNA Ligase I expression decreases after 72h of ZEB2 induction in A431-ZEB2 cells.

The repair of single strand breaks mainly takes place in the G1 phase where the cell is preparing for DNA replication (Branzei and Foiani, 2008). In addition, as shown previously, ZEB2 is thought to arrest cell cycle progression beyond G1 phase through direct repression of cyclin D1 (Mejlvang et al., 2007). Moreover, persistence of the single strand breaks was observed in cells expressing ZEB2 and this reduction is thought to be a result of the reduction in DNA Ligase I expression (Dr Gina Tse, Dr Eugene Tulchinsky lab, University of Leicester, unpublished data). So, does the repression of cyclin D1 correlate with the

repression of DNA Ligase I? Western blot data from A431-ZEB2 and DLD1-ZEB2 cells has confirmed the reduction in the expression of cyclin D1 and DNA Ligase I is correlated. Additionally, preliminary bioinformatics data from oncomine and several databases has shown correlation between cyclin D1 (*CCND1*) and DNA Ligase I (*LIG1*) in several squamous carcinomas (Dr Eugene Tulchinsky, unpublished data).

However, the mechanism by which EMT regulates LIG1 expression has not been elucidated yet. Bioinformatics' data has shown two areas of open chromatin in *LIG1*, at the promoter and intronic enhancer region (Figure 1.13).

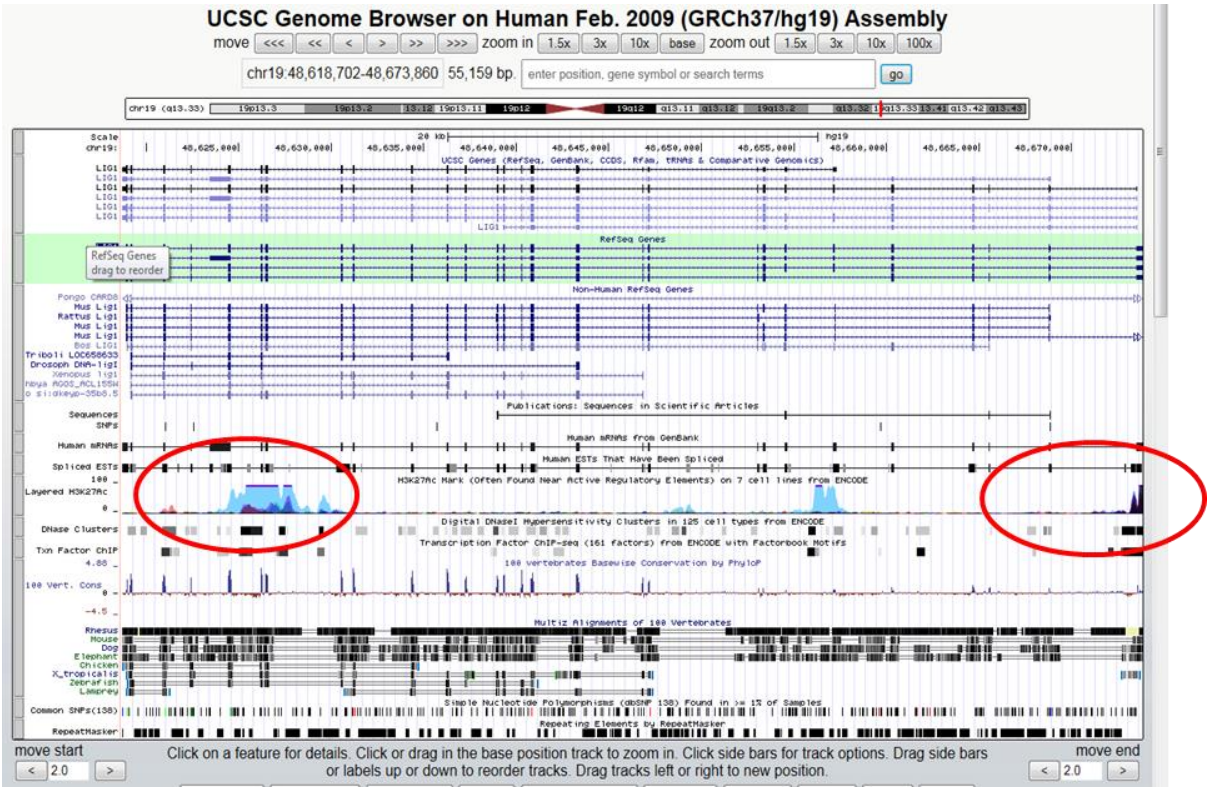


Figure 1.13: A map of the *LIG1* gene

showing the open chromatin areas of the promoter (right red circle) and first intron, a potential enhancer element (left red circle).

1.6 Project hypothesis and aims and objectives

Based on the previously discussed evidence, EMT causes a downregulation of DNA Ligase I expression, and this results in the decrease of its function in DNA repair and replication. This would result in accumulation of single strand breaks (SSB) in the G1 phase. These SSB will then be converted to double strand breaks (DSB) during DNA replication in S-phase as well as Okazaki fragments ligation problem will result in chromosomal aberration and genomic instability. So, we hypothesise that EMT is linked to genomic instability via the attenuation of DNA repair through the downregulation of the DNA Ligase I.

Chapter 2 : Materials and Methods

2.1 Materials

2.1.1 Cell culture materials

All of the cell lines used in this project were primarily purchased from the American Type of Culture Collection (ATCC, Rockville, Maryland), unless stated otherwise.

Table 2.1: cell lines

Cell line	Description	Culture conditions	Reference
A431-ZEB2	Human vulvar epidermoid carcinoma cells isolated from an 85 years old female. Prepared to conditionally express ZEB2 in the presence of DOX	DMEM, 10% v/v FBS and 1% v/v Pen/Strep	(Andersen et al., 2005)
A431-ZEB2-CyclinD1	Human vulvar epidermoid carcinoma cells isolated from an 85 years old female. Prepared to conditionally express ZEB2 and Cyclin D1 in the presence of DOX	DMEM, 10% v/v FBS and 1% v/v Pen/Strep	(Mejlvang et al., 2007)
DLD-1-ZEB2	Human colorectal adenocarcinoma isolated from an adult male. Prepared to conditionally express ZEB2 in the presence of DOX	RPMI, 10% v/v FBS and 1% v/v Pen/Strep	(Elenbaas et al., 2001)
HMLE-TWIST-ER	Immortalized human mammary epithelial cells that express Twist through activation of a fused modified oestrogen receptor (ER) by 4-OHT	DMEM (w/o L-glutamine and Phenol Red), 10% v/v FBS and 1% v/v Pen/Strep and 1% v/v GlutaMAx	(Mani et al., 2008)
MCF-7-ZEB1-GFP	Human breast adenocarcinoma cells isolated from pleural effusion of a 69 years old Caucasian female. Prepared to	DMEM, 10% v/v TET-minus or normal FBS and	Made by Dr Youssef Alghamdi at Dr Eugene Tulchinsky Lab

	conditionally express ZEB1 tagged to GFP in the presence of DOX	1% v/v Pen/Strep	
MCF-7-ZEB2-GFP	Human breast adenocarcinoma cells isolated from pleural effusion of a 69 years old Caucasian female. Prepared to conditionally express ZEB2 tagged to GFP in the presence of doxycycline	DMEM, 10% v/v TET-minus or normal FBS and 1% v/v Pen/Strep	Made by Dr Youssef Alghamdi at Dr Eugene Tulchinsky Lab
MDA-MB-231	Human breast adenocarcinoma isolated from the from metastatic site: pleural effusion of a 51 years Caucasian female	DMEM, 10% v/v FBS and 1% v/v Pen/Strep	Dr Emre Sayan group, University of Southampton
MDA-MB-231-shZEB1	Human breast adenocarcinoma isolated from the from metastatic site: pleural effusion of a 51 years Caucasian female transfected with small hairpin RNA to ZEB1	DMEM, 10% v/v FBS and 1% v/v Pen/Strep	Dr Emre Sayan group, University of Southampton

The MCF-7 cells used here have a wildtype p53 and Rb, ZEB proteins, DNA Ligases, p27Kip1, but have an inactive caspase 3 due to a frameshift deletion. They also have a splice site variation at the SMAD3 gene. Informations on the genetic background of the cell line were taken from the Broad Institute Cancer Cell Line Encyclopaedia (https://portals.broadinstitute.org/ccle/page?cell_line=MCF7_BREAST). They were transfected with the ZEB1 isoform number NM_001323676.

Table 2.2: cell culture reagents and materials

Reagent	Catalogue Number	Company
Dulbecco's Modified Eagle Media (DMEM) 4.5g/l glucose with L-Glutamine	12-604Q	Lonza, Belgium
Dulbecco's Modified Eagle Media (DMEM) 4.5g/l glucose without L-Glutamine and Phenol Red	BE12-917F	Lonza, Belgium
Roswell Park Memorial Institute Medium 1640 (RPMI) with Glutamine	BE12-702F	Lonza, Belgium
Heat Inactivated Foetal Bovine Serum (FBS)	EU-000-F	Sera Laboratories International Ltd, UK
Tet system approved FBS	631106	Clontech, France
Penicillin (5000 units) Streptomycin (5000 µg)	15070-063	Gibco by Life technologies
UV inactivated Trypsin/ Ethylenediaminetetraacetic acid (EDTA) (TE) 0.5%	15400-054	Gibco by Life technologies
Doxycycline hyclate (DOX)	D-9891-1G	Sigma-Aldrich, UK
Dimethyl sulfoxide (DMSO)	D5879	Sigma-Aldrich, UK
C-Chip haemocytometer	DHC-N01	Digital bio, UK
Cycloheximide (CHX)	C4859-1ML	Sigma-Aldrich, UK
Cryo.s™ cryogenic tubes	123623	Greiner Bio-one
CoolCell® Alcohol-Free Freezing container	13-900-856	Biocision / Fischer Scientific
GlutaMax (100x)	35050-038	Gibco by Life technologies
Geneticin® Selective Antibiotic (G418 Sulfate) (50 mg/mL)	10131-035	Gibco by Life technologies
Puromycin	A11138-03	Gibco by Life technologies
Mirus buffer, Ingenio® Electroporation solution	E7-0516	Geneflow, UK
Lipofectamine® 2000 Transfection Reagent	11668030	Invitrogen by Thermo Fisher Scientific, UK
Lipofectamine™ RNAiMAX Transfection Reagent	13778075	Invitrogen by Thermo Fisher Scientific, UK
Cdk4/6 Inhibitor IV (CINK4)- CAS 359886-84-3	219492-5MG	Merck Millipore, UK
(Z)-4-HYDROXYTAMOXIFEN (4-OHT)	H7904-5MG	Sigma-Aldrich, UK

2.1.2 Reagents and kits

Table 2.3: Reagents

Reagent	Catalogue Number	Company
2-mercaptoethnol (β-ME)	M7154	Sigma-Aldrich, UK
α-select chemically competent cells (Bronze)	BIO-85025	Bioline, UK
Acetic acid glacial	10304980	Fisher Scientific, UK
Agarose Hi-Res Standard	A4-0700	Geneflow, UK
Ammonium persulphate (APS)	A3678-100G	Sigma-Aldrich, UK
Ampicillin Na salt	A9518	Sigma-Aldrich, UK
Bovine serum albumin standard powder (Fraction V), 96% (BSA)	BPE9701-100	Fisher scientific, UK
Bromophenol Blue	B-8026-25G	Sigma-Aldrich, UK
Butan-1-ol 99%	B/4800/08	Thermo Fisher Scientific, UK
Chloroform	C2432-500ml	Sigma-Aldrich, UK
Colcemid	10295892001	Sigma-Aldrich, UK
4',6-diamidino-2-phenylindole (DAPI)	D9564	Sigma-Aldrich, UK
DEPC Treated Water	AM9906	Invitrogen, Ambion, UK
DPX Mountant for histology	6522	Sigma-Aldrich, UK
Ethanol Absolute (EtOH)	10437341	Fisher Scientific, UK
Ethidium bromide (EtBr)	46067	Fluka Biochemika
Giemsa stain	GS500	Sigma-Aldrich, UK
Glycine	G8898-1kg	Sigma-Aldrich, UK
Glycerol	G-5516	Sigma-Aldrich, UK
Hydrochloric acid (HCl)	H/1200/PB17	Thermo Fisher Scientific, UK
Ispopropanol	BP26181	Fisher Scientific, UK
Kanamycin	11815-024	Invitrogen, UK
KOD Hot start master mix	71842	Merck Millipore, UK
Lennox L Agar powder	22700-025	Invitrogen, UK
Luciferase Assay System 100 assays	E4030	Promega, UK
Methanol (MeOH)	M/3950/17	Fisher Scientific, UK
Miller's LB Broth Base	12795-027	Invitrogen, UK
N,N,N',N'-tetramethylethylenediamine (TEMED)	19281	Sigma-Aldrich, UK
o-Nitrophenyl β-D-galactopyranoside (ONPG)	73660-1G	Sigma-Aldrich, UK

PageRuler™ Plus Prestained Protein Ladder, 10 to 250 kDa	26620	Thermo Fisher Scientific, UK
HighRanger 1kb DNA Ladder (300bp-10000bp)	L3-0020/L3-0021	Geneflow, UK
FullRanger 100bp DNA Ladder (100bp-5000bp)	L3-0014/L3-0015	Geneflow, UK
PCR Ranger 100bp ladder (50bp - 1000bp)	L3-0004/L3-0005	Geneflow, UK
Phosphate buffered saline (PBS) tablets, pH 7.3	BR0014G	Oxoid Ltd, UK
Ponceau S	78376-100G	Sigma-Aldrich, UK
Propidium Iodide	P4170 - 100mg	Sigma-Aldrich, UK
30% Protogel Acrylamide	A2-0072	GeneFlow, UK
Sodium dodecyl sulphate (SDS), 20% solution	10607443	Fisher Scientific, UK
S.O.C media	S1797	Sigma-Aldrich, UK
Sodium Chloride	S7653-1kg	Sigma-Aldrich, UK
Tris-acetate-EDTA Buffer (TAE) (50X)	B49	Thermo Fisher Scientific, UK
Tris-borate-EDTA Buffer (TBE) (10X)	B52	Thermo Fisher Scientific, UK
TESCO Dried Skimmed Milk	-	TESCO, UK
Tris Base	BPE 152-1	Thermo Fisher Scientific, UK
TRIzol® reagent	15596-026	Invitrogen, UK
Tween-20	P5927	Sigma-Aldrich, UK

Table 2.4: Commercial kits

Kit	**Catalogue Number**	**Company**
AllPrep DNA/RNA Mini Kit	80204	Qiagen, UK
NucleoSpin® plasmid	NZ-74058850	Macherey-Nagel, Fisher Scientific, UK
NucleoBond Xtra Maxi Plus	740414.5	Macherey-Nagel, Fisher Scientific, UK
Pierce BCA Protein Assay Kit	23227	Thermo Scientific, UK
Pierce ECL Western Blotting Substrate	32106	Thermo Scientific, UK
Pierce Supersignal West Dura Extended Duration Substrate	34076	Thermo Scientific, UK
Wizard® SV Gel and PCR Clean-Up System	A9281	Promega, UK

2.1.3 General equipment

Table 2.5: General equipment

Equipment	Catalogue number	Company
CL xposure film 180mm x 240mm	34089	Thermo Scientific, UK
Curix 60 X-ray film processor		AGFA Healthcare, UK
Electroporation Cuvettes 4mm	E6-0060	GeneFlow, UK
GenePulser Xcell electroporator	1652661	Bio-Rad Laboratories, UK
G:BOX gel doc		Syngene
GeneAmp PCR system 2400		Perkin Elmer, USA
Immobilon-P polyvinylidene difluoride (PVDF) membrane	IPVH 00010	Merck Millipore, UK
Microscope slides, twin frosted 76 mm x 26 mm, 1.00 to 1.2 mm thick	11572203	Fisher Scientific, UK

2.1.4 Primers, plasmids and siRNAs.

qPCR primers were designed by NCBI/primer-BLAST. All primers were purchased from Sigma. Pelleted primers were resuspended in DEPC treated water to a 100 µM concentration stock solution that was then diluted to a working solution of 10µM.

Table 2.6: Primers

Target	Forward	Reverse	Product size (kb)	Tm (°C)	No. of Cycles	Use
LIG1 promoter	CACCacgcgtCGAAGAAGCGGCTGAACTCGGC	CACCagatctCCAAGCATTCGGCGCACCCGCC	1.7	65	30	cloning
LIG1 enhancer	CAggatccGTGAGCCCCCAGAAGGAGAGAAG	CCAgtcgacCTGACGATAGACAGAACGGTCAGATG	2.5	65	30	cloning
LIG1 promoter	CACCacgcgtACCTCCCACAATGCCCCGCGCC	CACCagatctCCAAGCATTCGGCGCACCCGCC	0.5	65	30	cloning
LIG1 promoter Swap	CACCagatctCGAAGAAGCGGCTGAACTCGGC	CACCacgcgtCCAAGCATTCGGCGCACCCGCC	1.7	65	30	cloning
CDKN1B promoter (1)	CACCgagctcAGAGCCTAAGGATAATGTCTGGCATAGA	CCGGTACCCATCATCTTGGTTTGAGCCAAAGTTA	~2	65	35	cloning
CDKN1B promoter (2)	TATGAGATGAGGTAGGCACACAAAGTGGACAAG	CACCaagcttACAGAGGAGGAGATCCATTGGTTGCG	~1.5	65	35	cloning
3'UTR of CDKN1B	CACCgagctcCCTCAGAAGACGTCAAACGTAAAC	CACCaagcttGAGGGGAAAACCTATTCATACCC	1.4	64	25	cloning
CDKN1B	TAATTGGGGCTCCGGCTAAC	GAAGAATCGTCGGTTGCAGGT				PCR/qPCR
GAPDH	GGCTGAGAACGGGAAGCTTGTCAT	TCTTCACCACCATGGAGAAGGCTG				PCR/qPCR

Table 2.7: Plasmids

Plasmid name	Plasmid	Insert	Use	Catalogue number	Supplier/ Made by
pGL3-Basic	pGL3-Basic vector	-	Luciferase assay negative control	E1751	Promega, UK
pGL3-Promoter	pGL3-Promoter vector	-	Luciferase assay positive control as it has SV40 promoter upstream of the luciferase gene	E1761	Promega, UK
pGL3-*LIG1* pr-1727	pGL3-Basic vector	Full length human WT *LIG1* promoter	To assess the activity of *LIG1* promoter by luciferase assay	-	Gina Tse
pGL3-*LIG1* pr- 500	pGL3-Basic vector	First 500bp of WT *LIG1* promoter	To assess the activity of *LIG1* promoter by luciferase assay	-	Noura Alibrahim
pGL3-LIG1 pr swap-1727 (pGL3-C19orf68-1727)	pGL3-Basic vector	Full length human WT *LIG1* promoter in the opposite direction (C19orf68 promoter)	To assess the activity of the opposite direction of *LIG1* promoter by luciferase assay	-	Noura Alibrahim
pGL3-Promoter-*LIG1* enh-2532	pGL3-Promoter vector	Full length Human WT *LIG1* enhancer	To assess the activity of *LIG1* enhancer by luciferase assay	-	Noura Alibrahim
pGL3- *LIG1* pr-1727-*LIG1* enh-2532	pGL3-Promoter-*LIG1* enh-2532	Full length human WT *LIG1* promoter	To assess the activity of *LIG1* promoter when coupled to the enhancer by luciferase assay	-	Noura Alibrahim
pCMV β-gal	pCMV-SPORT1 Invitrogen 10586-014	β-galactosidase	Transfection efficiency internal control for luciferase assays	-	In lab
pGL3-*CDKN1B* promoter-3500	pGL3-Basic vector	Full length human WT *CDKN1B* promoter	To assess the activity of *CDKN1B* promoter by luciferase assay	-	Noura Alibrahim
pMIR-REPORT	pMIR-REPORT vector Promega	-	negative control for miRNA target validation by luciferase assay	-	In lab
pMirTarget-3'UTR - *CDKN1B*	pMirTarget vector	Human WT 3'UTR of *CDKN1B*	miRNA target validation by luciferase assay	SC214055	OriGene Tech, USA
pMIR-*CDKN1B*-954	pMIR-REPORT vector	Human WT 3'UTR of *CDKN1B*	To validate *CDKN1B* as miRNA target by luciferase assay	-	Noura Alibrahim
pMIR-*CDKN1B*-1400	pMIR-REPORT vector	Human WT 3'UTR of *CDKN1B*	To validate CDKN1B as miRNA target by luciferase assay	-	Noura Alibrahim

Lyophilised siRNA oligonucleotides were resuspended in 100µl of DEPC treated water to a final concentration of 50nmol/µl. for each transfection, 3µl were used.

Table 2.8: siRNAs

Gene	Sense	Antisense	Source	siRNA ID/Catalogue no.
Negative control	AUGAACGUGAAUUGCUCAA[dT][dT]	UUGAGCAAUUCACGUUCAU[dT][dT]	Sigma-Aldrich, UK	4457287
CCND1	GGAGCAUUUUGAUACCAGAtt	UCUGGUAUCAAAAUGCUCCgg	Ambion, USA	S229/4390825
LIG1	CAAGAAAGAGGGUAAAGCAtt	UGCUUUACCCUCUUUCUUGgg	Ambion, USA	S8173/4390825
CDKN1B	CAAACGUGCGAGUGUCUAA		Dharmacon, GE healthcare UK	L-003472-00-0005

2.1.5 Restriction and ligation enzymes

Table 2.9: Enzymes.

Enzyme	Catalogue number	Company
Antarctic phosphatase (5,000 units/ml)	M0289S	New England Biolabs, UK
BamHI (20,000 units/ml)	R0136S	New England Biolabs, UK
BglII (10,000 units/ml)	R0144S	New England Biolabs, UK
HindIII (20,000 units/ml)	R0104S	New England Biolabs, UK
KpnI (10,000 units/ml)	R0142S	New England Biolabs, UK
SalI (20,000 units/ml)	R0138S	New England Biolabs, UK
XhoI (20,000 units/ml)	R0146S	New England Biolabs, UK
T4 DNA Ligase (5U/µl)	15224-041	Invitrogen, UK

2.1.6 Western blotting antibodies

Table 2.10: Primary antibodies

Protein	Migration in SDS/PAGE (kDa)	Antibody/clone no.	Catalogue No.	Source	Dilution
α-tubulin	55	Mouse monoclonal (B-5-1-2)	T5168	Sigma-Aldrich, UK	1:40000
Cyclin D1	35	Rabbit monoclonal (2978)	92G2	Cell signalling technology, UK	1:250
DNA Ligase I	~130	Mouse monoclonal (C-5)	sc-271678	Santa Cruz Biotech, Germany	1:1000
DNA Ligase III	~103	mouse monoclonal (7)	sc-135883	Santa Cruz Biotech, Germany	1:200
DNA Ligase IV	100	Rabbit monoclonal (D5N5N)	14649	Cell signalling technology, UK	1:1000
E-cadherin	~120	Mouse monoclonal (36)	610181	BD Transduction Laboratories, UK	1:2000
p27Kip1	27	Rabbit polyclonal (M-197)	sc-776	Santa Cruz Biotech, Germany	1:2000
phospho-H2A.X (Ser139)	15	Mouse monoclonal (JBW301)	05-636	Merck Millipore, UK	1:100
Pirh2	30	Rabbit monoclonal (EPR14980)	Ab189247	Abcam, Cambridge, UK	1:1000
Rb	~105	Mouse monoclonal (LM95.1)	OP66-100UG	EMD Millipore, Germany	1:1000
ZEB1	250	Rabbit polyclonal (#H1513)	sc-25388	Santa Cruz Biotech, Germany	1:1000
ZEB2	250	Rabbit polyclonal	Made in house, Dr Eugene Tulchinsky (University of Leicester, UK)		1:3000

Table 2.11: Secondary antibodies

Antibody	Type	Catalogue No.	Source	Catalogue No.
Anti-Rabbit Immunoglobulins/HRP	Goat polyclonal	P0488	DAKO Ltd, UK	P0488
Anti-Mouse Immunoglobulins/HRP	Goat polyclonal	P0447	DAKO Ltd, UK	P0447

2.2 Methods

2.2.1 Cell line routine maintenance and splitting

Cell lines monolayers were grown according to respective culturing conditions (Table 2.1) and maintained in a 5% CO_2 incubator at 37°C and 100% humidity. Cells were split at around 80% confluency. Before splitting, the cells were washed twice with PBS and detached by 2x TE for 5-15 minutes at 37°C. After that, TE was inactivated by ×3 amount of the respective media and the cells were spin down for 5 min at 1000rpm. Cells were then resuspended in fresh media and then split or seeded as required. A disposable C-chip™ haemocytometer was used for cell counting according to manufacturer protocol.

2.2.2 Cell lines freezing and thawing

For freezing, cells were collected as described in 2.2.1 and resuspended in 10ml of freezing media (70% v/v cell culture media, 20% v/v FBS and 10% v/v DMSO). Cells were then aliquoted into the freezing vials (Cryo.s™) that are placed in the CoolCell® freezing container at -80°C overnight. Cryo.s™ vials are then transferred to liquid nitrogen for long term storage.

For thawing, after removing from liquid nitrogen, cells are immediately placed in a water bath at 37°C. Once thawed, cells are mixed with 10ml of respective cell culture media and centrifuged for 5 min at 1000rpm. Cells are then resuspended in 5ml of respective media before transferring to a 25cm^2 flask.

2.2.3 Cell lines transient transfection

Cell lines were transiently transfected with siRNAs or plasmids for western blotting or luciferase assays. Transfections were performed by electroporation. In electroporation, cellular plasma membranes are permeablised through the application of an electrical field across them, which allows the entry of the test materials (i.e. plasmids or siRNA) into the cell.

The cells were initially trypsinised, centrifuged for 5min at 1000 rpm, resuspended in 1ml of PBS and counted. Then, the cells were aliquoted into 1.5 ml microcentrifuge tubes. Approximately 0.5-2 × 10^6 cells were used for each transfection depending on the plate or dish to be used. The cells were centrifuged again for 5min at 1000rpm then resuspended in 60µl of the Ingenio® Electroporation buffer. The cells were then mixed with the DNA and/or siRNA to

be transfected followed by transferring the mixture to 4mm electroporation cuvettes. The cuvette then was placed in the GenePulser Xcell electroporator and electroporated at a voltage of 250 V and a capacity of 250 µF. The cells were then immediately transferred to 6-well plates, 60mm or 100mm dishes containing pre-warmed treated media as appropriate. The plates or dishes were then returned to the incubator and the cells were collected after 72h as required.

2.2.4 Cell lines chemical treatments

2.2.4.1 Doxycycline treatment

In order to express ZEB1/ZEB2 (i.e. induce EMT), cells were treated with 2 µg/ml of DOX. The cells were treated with DOX from the day of seeding for 72h (MCF-7 and A431 cell lines) or 120h (DLD1 cell line) before harvesting.

2.2.4.2 4-OHT treatment

To express Twist-ER and induce EMT, HMLE cells were treated with 40ng/ml of 4-OHT. 4-OHT was added to the cells from the day of seeding for 12 days with changing the media containing 4-OHT every 3 days.

2.2.4.3 Cdk4/6 inhibitor IV Treatment

To investigate how EMT regulates DNA ligase I expression, and whether this regulation is linked to the cell cycle, cells were treated with the Cdk 4/6 inhibitor IV (CINK4) (Merck Millipore, UK). CINK4 inhibits the enzymatic activity of CdK4 by acting as an ATP competitive and reversible inhibitor to cdk4/6-complexed cyclin D1. The cells were treated with CINK4 one day after seeding for 72 hr before harvesting. Cells were seeded at 0.5×10^6 density in 60mm dishes. A titration experiment was performed initially to identify the most effective but non-toxic concentration of the inhibitor. The concentrations used were 10µM, 5µM, 4µM and 0.5µM.

2.2.4.4 Cycloheximide treatment

CHX blocking was performed to determine the stability of p27Kip1 protein through determining its half-life. CHX inhibits protein biosynthesis by blocking translational elongation. It does so by binding to the 60S ribosomal unit's E-site and blocking deacetylated tRNA binding (Schneider-Poetsch et al., 2010).

Approximately 0.2×10^6 cells were seeded into 6 well plates with or without DOX for 72h before CHX treatment. DOX treated cells were washed twice with PBS

before adding CHX. Cells were treated with 20µg/ml CHX for 0, 2, 4, 6, 8 or 16h before lysing. The 0 time point was treated with DMSO as a negative control.

2.2.5 Analysis of cell cycle distribution

Cell cycle distribution analysis was performed using Propidium Iodide (PI) and flow cytometry whereby DNA content (hence cell cycle phase) is measured. This is done by quantifying the amount of fluorescence emitted by dye present in the cell which is directly proportional to cellular DNA content.

Cells were seeded and treated with Dox as required. When harvesting, at ~80% confluency, cells were washed once with PBS then detached with 2×TE. TE was inactivated with normal media and the cells were centrifuged for 5min at 1000rpm. Cells were resuspended in PBS and centrifuged again before resuspending in 200µl of PBS. Then, to fix and permeablise the cells so that PI and RNase can enter the cell, 2ml of ice cold 70% v/v EtOH was added dropwise to the cells while mixing on vortex. Cells were then allowed to fix for 90min on ice. After that, cells were centrifuged for 10min at 600×g before resuspending in 800µl of PBS. Then, 100µg/ml of RNase and 50µg/ml of PI was added to the cells before mixing on a shaker. Cells were then incubated overnight at +4°C and protected from light. Next day, cells were analysed in BD FACS Aria II and analysed on FACSDiva software Version 6.1.3 (Mejlvang et al., 2007).

2.2.6 MCF-7 growth curve

To analyse the effect of p27 knockdown on EMT induced MCF-7 cells, a growth curve was performed. Here, 0.8×10^6 cells were transfected (section 2.2.3) with either siCONTROL or siCDKN1B (table 2.8) and seeded in 6cm dishes to grow overnight. Next day, cells were seeded in triplicates in 6-well plates at 3×10^4 densities with/without Dox. Cells were counted using C-chipTM haemocytometer at 24h, 48h, 72h and 96h post Dox treatment. The average of cells number for each time point was calculated for each time point for both siCONTROL and siCDKN1B transfected cells. A graph was generated using Microsoft Excel showing mean and standard deviation (SD) for both cell groups.

2.2.7 Nucleic Acid preparations

2.2.7.1 RNA Extraction

RNA extraction was performed using a combined protocol of the TRI Reagent® procedure (Chomczynski and Sacchi, 1987) and the RNeasy® Plus Mini Kit. Cells were washed twice with PBS before adding 1ml of the TRI Reagent®, which is composed of phenol and guanidine thiocyanate to help in dissolving proteins and nucleic acids. The cells were then lifted from the plates using a scraper and transferred into a 1.5ml tube. The cells were then left to incubate for 5min at RT before adding 200µl of chloroform and mixing samples very well to ensure that chloroform and the TRI Reagent® were combined well. Samples were then centrifuged at 12000g at 4°C for 15min. After centrifugation, 3 layers of solution were produced where the lower layer collects proteins, the middle containing DNA and the upper containing the RNA. Then, the upper aqueous phase was carefully collected, and the purification procedure continued using the RNeasy® Plus Mini Kit and following manufacturer procedure. At the end, RNA was eluted using RNase free water and stored at -80°C.

2.2.7.2 RNA/DNA concentration measurement

NanoDrop ND-1000 spectrophotometer was used to determine the concentration and yield of nucleic acids. Initially, the pedestal and lid of the equipment were cleaned with the instrument cleaner and the program was initialised with 1.2µl of DEPC treated water. Then, it was blanked with the same solution used to elute the nucleic acid. 1.2µl of samples were also loaded on the pedestal and their concentrations were measured in duplicates. The purity of the sample was defined based on the absorbance ratios at 260nm and 280nm (260/280 ratios) with an optimal ratio of 1.8 for RNA and 2.1 for DNA.

2.2.7.3 cDNA synthesis

cDNA synthesis was performed using the RevertAid^{TM} First Strand cDNA synthesis kit. For each cDNA synthesis reaction, 1µg of RNA was used before making the solution up to 11µl with nuclease free water. Then, 1µl of random primers was added to the diluted RNA and mixed well. The sample was then centrifuged briefly and incubated for 5min at 65°C. They were then centrifuged again and placed on ice. After that, the rest of the reaction was added to the RNA/ primers mix as shown in table 2.12 and in the same order.

Table 2.12: cDNA synthesis mix

Reagent	Volume (µl)
5 × Reaction buffer	4
RiboLock RNase inhibitor	1
10mM dNTP mix	2
RevertAid Reverse Transcriptase	1
Total	20

The reagents were centrifuged briefly again after mixing and the reaction was initiated by incubating at 25°C for 15min. After that, the reaction was incubated for 45min at 42°C followed by reaction termination at 70°C for 5min. The samples were then stored at -20°C until needed.

In parallel to the cDNA synthesis reaction, positive (GAPDH) and negative (minus the RTase) controls were also prepared to monitor even RNA content and quality of the sample and the freshly synthesised cDNA. The GAPDH primers (10µM) were included in the kit and the RT-PCR reaction was performed for each sample as shown in table 2.13.

Table 2.13: cDNA positive control reaction mix

Reagent	Volume (µl)
2 × Taq BioMix red	10
GAPGH forward primer	0.5
GAPGH reverse primer	0.5
Nuclease free water	7
cDNA	2
Total	20

Reactions were then mixed gently, centrifuged briefly then transformed to the thermal cycler and were run at the following conditions:

- Initial denaturation: 5min at 95°C
- 25 cycles (denaturation 30sec at 95°C, annealing for 30sec at 60°C and extension for 30sec at 72°C)
- Termination: 5min at 72°C

After that, the PCR products were run in 1% agarose/TAE gel that contains ethidium bromide in a horizontal Bio-Rad electrophoresis tank for 30min at 90V. Gel was then visualised in a G:BOX Gel Doc under UV to confirm the presence of GAPDH bands.

2.2.7.4 qPCR

qPCR can be used to determine the relative change in the expression level of a gene of interest upon change in the state of the cell or for example a change in culturing conditions (Schmittgen and Livak, 2008). qPCR was performed to analyse *CDKN1B* level of expression upon EMT induction by ZEB1 in MCF7 cells.

CDKN1B primers were diluted in a 1:10 ratio and the cDNA was diluted in 1:20 prior to making the stock solutions (table 2.14 and 2.15) on ice and in clean DNase free tubes. Before using, tubes were gently mixed and centrifuged briefly. As a housekeeping gene, *GAPDH* was used and a primer mix stock solution was prepared as described for *CDKN1B* primers.

Table 2.14: SYBR green and primers master mix

Reagent	Volume (µl)
SYBR green	10
Forward primer	0.5
Reverse primer	0.5
Total	11

Table 2.15: diluted cDNA stock solution

Reagent	Volume (µl)
Diluted cDNA	4
Nuclease-free water	5
Total	9

On ice, a 96-well plate qPCR was placed and 11 µl from the SYBR green and primers stock solution was added to the wells followed by 9 µl of the cDNA stock solution. In addition, to monitor contaminations, a negative control was performed where water was added instead of cDNA. Then, MicroAmp® optical film was used to seal the plate before spinning it down briefly and loading it onto the StepOnePlus™ Real- Time PCR system. Once the test was complete, data were analysed using the $2^{-\Delta\Delta C}T$ method.

2.2.7.5 Determination of relative change in a gene expression level

The $2^{-\Delta\Delta C}T$ method was utilised to analyse the qPCR data for *CDKN1B* level of expression. This method is used to determine the relative change in a gene of interest expression level in relation to a reference group, a control group for instance. This system depends on accepting that the efficiency of both

amplifications is equal and that primers efficiency is maximal (Livak and Schmittgen, 2001). The cycle threshold (C_T) shows the cycle number where the amplified target amount hits a set threshold. This threshold was set at 0.1. The C_T value is inversely proportional to the amount of nucleic acid present in the sample. The analysis method consists of the follow:

Δ= the change

A= C_T of the gene of interest in the control sample

B= C_T of the gene of interest in the treatment sample

hc= the average of C_T values of the housekeeping gene in the control sample

ht= the average of C_T values of the housekeeping gene in the treatment sample

Firstly, the C_T of the gene of interest is normalised to that of the housekeeping gene ($\Delta C_{T\ A\text{-}hc}$ or $\Delta C_{T\ B\text{-}ht}$). After that, the difference in C_t between the two samples is determined ($\Delta\Delta C_T = \Delta C_{T\ B\text{-}ht} - \Delta C_{T\ A\text{-}hc}$). Eventually, the fold difference between the two samples is calculated using the equation $2^{-\Delta\Delta C}T$.

2.2.8 Cloning via PCR

Initially, an insert of interest is amplified using PCR, from either a donor plasmid or human DNA with concomitant incorporation of restriction enzymes specific to the recipient vector. After that, the insert and vector are digested with the same set of enzymes and purified. Eventually, the insert and recipient vector are ligated creating a new plasmid.

2.2.8.1 Target amplification

Targets were amplified using target specific primers sets (table 2.6). PCRs were done in 50µl reactions that consisted of 1µl of each of the forward and reverse primers (10pmol/µl), 1µl of either the whole human DNA or from the donor plasmid (at 200ng concentration), 22µl nuclease-free DEPC-treated water and 25µl KOD Hot Start Master Mix. The PCR machine was programmed to pre-heat for 2min at 95°C, then two cycles (denaturation for 20s at 95°C, annealing for 30s at primers specific Tm and extension for 1min at 70°C) to incorporate the restriction sites of the enzyme to be used. Then, 30 cycles (denaturation for 20s at 95°C, annealing and extension for 1.5 min at 70°C) and a final extension of 7 min at 70°C. Then, 5µl of the products was run in 1% w/v agarose/TAE gel to confirm that the presence of the insert and was run as described previously (section 2.2.5.3). After confirmation, the PCR products were column purified using Wizard® SV Gel and PCR Clean-Up System according to the manufacturer procedure. The DNA was eluted in 25µl of nuclease-free DEPC-treated water.

2.2.8.2 Target and vector preparation

After purifying the amplified targets, the targets were cut using the appropriate restriction enzymes, as a preparation for ligation. In a 20or 30µl reaction mix: the whole amount of the target was mixed with 1µl of each of the appropriate restriction enzymes, 10x appropriate buffers and xµl nuclease-free DEPC-treated water as needed. The reactions were incubated at 37°C for 2h. Simultaneously, in a 15µl reaction: 1-2µg of the appropriate vector were cut with 1µl of each of the enzymes used for the target, 1.5µl of 10x appropriate buffer and xµl nuclease-free DEPC-treated water as needed. To increase the accuracy of the ligation, the 5'-end of the vector was dephosphorylated using antarctic phosphatase in a 30µl reaction that contains: all of the digested vector, 1µl of the enzyme, 3µl of the 10x antarctic phosphatase buffer (50 mM Bis-Tris-Propane-HCl, 1mM $MgCl_2$ and 0.1 mM $ZnCl_2$) and xµl nuclease-free DEPC-treated water as needed. The reaction was incubated at 37°C for 30min followed by 5min at 70°C to inactivate the phosphatase. The target and vector were then column purified with Wizard® SV Gel and PCR Clean-Up System and were eluted in 25µl nuclease-free water. The appropriate bands were then confirmed in a 1% w/v agarose/TAE gel.

2.2.8.3 Target and vector Ligation

The ligation reactions were performed using T4 DNA Ligase. In a 20μl reaction, 2-4μl of the digested vector were mixed with variable amounts of the digested target (0.5, 1, 2, 3, 8μl or 0 in negative controls), 2μl of 10x T4 DNA Ligase buffer, 0.5 μl of the Ligase and xμl nuclease-free DEPC-treated water as needed. The reaction was incubated for 1h at RT.

2.2.8.4 Transformation of E.Coli and plasmid DNA production

Plasmids produced from the ligation were then transformed into α-select competent cells of Bronze efficiency. Initially, the competent cells were thawed on ice, as they are stored at -80°C. While the cells thaw, 1.8μl of the DNA plasmids and 1μl of the positive control pUC19 (provided with the cells) were aliquoted in 1.5 microcentrifuge tubes and left on ice to chill. After thawing, cells were mixed by gently pipetting up and down and 30μl of cells was added to each of the pre-chilled tubes and mixed gently. The tubes were then incubated on ice for 30min before being heat-shocked for 30sec in a 42°C water bath. This heat shock would permeablised the bacteria so they take up the plasmid. Cells were then re-incubated on ice for 2min. Then, 300μl of SOC medium was added to each tube and shaken (~300rpm) for 1h at 37°C. Finally, in a bacterial hood, the whole content of the tubes was spread on LB agar plates and incubated overnight at 37°C. To prepare the agar plates, previously prepared and autoclaved agar (12.5g of Lennox L Agar powder (10g SELECT peptone-140, 5g SELECT yeast extract, 5g NaCl and 12g SELECT agar/L) made up to 400ml with deionised water) was melted completely and then cooled down to 55°C before adding the antibiotic Ampicillin (at 100μg/ml). Finally, the agar was poured into 100mm petri-dishes and left to set at room temperature.

2.2.8.5 Identification and isolation of plasmid DNA

2.2.8.5.1 Small scale bacterial cultures

To prepare the small cultures (starter cultures), colonies were directly picked from the agar plates and added to tubes containing 3ml of Luria Broth (LB) broth (20g of Miller's LB Broth Base powder (10g SELECT peptone-140, 5g SELECT yeast extract, 10g NaCl/L) made up to 800ml deionised water) with 100 μg/ml ampicillin. The cultures were then incubated overnight in a shaking incubator at 37°C. Next day, 1.5 ml of the cultures was used to isolate the plasmid DNA using

the NucleoSpin® Plasmid kit following the manufacturer protocol for isolating high-copy plasmid DNA from *E.coli*. The plasmid DNA was eluted in 50µl of buffer AE (5 mM Tris/HCl, pH 8.5). Then, 1µl of the plasmid DNA from each culture was run in a 1% agarose gel alongside the negative control. Plasmids that were heavier than the negative control were further analysed, through restriction enzymes and sequencing, to confirm the recombinant plasmid.

2.2.8.5.2 Large scale bacterial cultures

Once recombinant plasmid DNA was confirmed, 300µl of the plasmids starter culture was added to 300ml of LB media containing ampicillin at the previously mentioned concentration. The culture was then incubated overnight in a shaking incubator at 37°C. Next day, the plasmid DNA was isolated using the NucleoBond® Xtra Maxi Plus kit following the manufacturer protocol for high-copy plasmid purification. The DNA was eluted in 400µl of Buffer Tris (5 mM Tris/HCl, pH 8.5). The concentration and purity of the plasmid DNA was then measured as described in section 2.2.5.2. Plasmid DNA with a concentration <1µg were re-concentrated by adding 1/20th the volume of 5M NaCl and 2.5 volume 95% ethanol. These were then centrifuged at high speed (13200rpm) for 10 min. Then, the supernatants were discarded and the tubes centrifuged again briefly to collect the ethanol down, which was discarded carefully. Then, the pellets were resuspeded in 100-150µl of nuclease-free DEPC-treated water. Finally, the plasmid DNA was confirmed again with restriction enzymes and visualised in 1% agarose/TAE gel. Then, the plasmids were sent for sequencing for the final confirmation of the recombinant plasmids.

2.2.8.5 Plasmid DNA sequencing

For sequencing, recombinant plasmids were sent to GATC Biotech Ltd (London, UK). Plasmids were diluted to 80-100ng/µl and was mixed separately with 5µM of each of the forward and reverse primers in 1.5ml eppendorfs and were sent to the company. Sequencing data were then blasted in NCBI data base and aligned with human genes to confirm the success of the cloning.

2.2.9 Protein analysis (western blotting)

Here, the proteins are separated according to their molecular weight by electrophoresis as they move from negative to positive charge at a stable voltage

over time (Mahmood and Yang, 2012). The proteins are stacked at a pH of 6.8 with a low concentration of polyacrylamide, to form distinct and thin bands, and are resolved at a pH of 8.8 with a higher concentration of polyacrylamide. All western blotting experiments were repeated at least twice and representative images are shown in the results.

2.2.9.1 Protein isolation

At approximately 80% confluency, cells were washed twice with PBS and then an x amount of 1x Laemmli buffer (50mM Tris-HCl pH 6.8, 20% v/v SDS and 10% v/v glycerol) was added to the flasks/dishes. The amount of Laemmli buffer to be added depends on the number of cells in the flasks/dishes. For 6-well plates, 6cm dishes and 10cm dishes, the amount of Laemmli buffer was 80µl, 150µl and 300µl, respectively. Cells were then detached from the base of the flask/dish by scrapping and the lysis buffer containing the cells was then pipetted into 1.5ml microcentrifuge tube. The samples were then homogenised by sonication for 10-30 seconds using a Soniprep 150 sonicator before determining the concentration of the proteins.

2.2.9.2 Determination of protein concentration

The concentration of the proteins in the samples was determined using the Pierce™ BCA protein assay kit, which is a colourometric based procedure that depends on bicinchoninic acid (BCA) at which a purple colour develops following two molecules of BCA chelation with a cuprous ion (Wiechelman et al., 1988). The quantitation procedure was performed according to manufacturer's instructions.

2.2.9.3 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Gels were casted in 1.5mm vertical glass plates and the electrophoresis was performed using the mini-protein tetra electrophoresis system. Resolving gels of various polyacrylamide percentages were prepared, depending on the size of the protein to be analysed. The gels consisted of 0.375M Tris (pH 8.8), 6-12% v/v of 30% polyacrylamide (37.5:1 acrylamide:bisacrylamide) and 1% v/v of (10% v/v SDS). Then, to initiate polyacrylamide polymerisation, 1% v/v of (10% w/v ammonium persulfate (APS)) and 0.1% v/v of N,N,N,N-

Tetramethylethylenediamine (TEMED) were added to the mixture immediately before pouring into the casting plates. After pouring, water saturated butanol was pipetted on top of the gel to avoid gel evaporation whilst setting and to remove any bubbles from the gel interface. The gels then were left to set for 30-40min. The butanol was then washed away with deionised water before pouring the 5% stacking gel that contained 1M Tris-HCl, pH 6.8, 5% v/v of 30% polyacrylamide, 10% v/v SDS , 1% v/v of (10% w/v APS) and 0.1% v/v of N,N,N,N-Tetramethylethylenediamine (TEMED) to initiate the polymerisation of the gel. The stacking gel was poured on top of the resolving gel and a 15 or 10 wells comb was place and the gel was left for another 10-20 min to set.

Whilst the gels were prepared, 1µg/µl protein samples were prepared and mixed with 5x protein loading buffer before boiling for 5 min at 95°C. The samples were then left to cool to room temperature before loading. The casted gels were then assembled in the electrophoresis tanks and the tanks were filled with 1x running buffer (25 mM Tris, 190 mM glycine and 0.1% (w/v) SDS). The protein samples were then loaded into the casted wells, 20µg/well, along with 5µl of the PageRuler Plus^{TM} Prestained protein ladder. Gels were then run using Bio-Rad PowerPac HC for 60-70min at 170V.

2.2.9.4 Protein transfer from gel to Polyvinylidene difluoride (PVDF) membrane

Following electrophoresis, the proteins within the gel were transferred into the Immobilon-P PVDF membrane and utilising the wet electrophoresis transfer system, Trans-Blot system. In this system, a constant voltage is applied perpendicular to the gel-membrane so, the proteins move from the gel to the membrane (Mahmood and Yang, 2012). Initially, for activation, the membrane was immersed in 100% v/v methanol for 1min before equilibration in 1x transfer buffer (25mM Tris, 190mM glycine and 20% (v/v) methanol) for few minutes. The transfer cassette, the sponge pads, filter papers, and the gels were immersed in the transfer buffer for few minutes prior to assembling the transfer cassette. The transfer cassette was then assembled in the following order: Black side (negative); sponge pad, 2x filter paper; gel; PVDF membrane; 2x filter paper; sponge pad; clear side (positive). Bubbles were carefully removed from the sandwich by rolling over before closing the cassette. The cassette was then

properly placed in the transfer tank which was filled with 1x transfer buffer. Transfer was performed for 16 h at 25 V.

2.2.9.5 Staining and immunodetection of proteins

Once transfer finished; the membranes were immersed briefly in Ponsceau S stain (0.1% (w/v) Ponsceau S in 5% (v/v) acetic acid) to ensure the success of the transfer and facilitate cutting the membrane if necessary. Excess staining was washed away using 1xTBS-T (50mM Tris pH 7.65, 150mM NaCl, and 0.1% (v/v) Tween® 20). The membranes were then blocked in 5% TBST-milk (1x TBST, 5% (w/v) fat free milk) for 1 hr on a shaker at room temperature. After that, the membranes were washed 3 times in TBST for 5min each with shaking. The membranes were then placed in the primary antibodies at the required concentration (table 2.10) in 5% TBST-BSA (1xTBST, 5% (w/v) BSA) and incubated rotating for 1 hr at room temperature. After that, the membranes were washed 4 times in TBST for 10min each with shaking to ensure reduce background by removing excessive unbound antibody. Then, the membranes were placed in the horseradish peroxide (HRP) conjugated secondary antibody (table 2.11) in 5% TBST-milk and incubated rotating for 1 hr at room temperature. The membranes were washed again as after the primary antibody. To detect the protein bands, the activity of the peroxidase was detected using either the luminal-based Pierce® enhanced chemiluminescence (ECL) western blotting substrate or SuperSignal™ West Dura Extended Duration Substrate (Thermo Scientific, UK). The theory behind this detection method is that the substrate gets oxidised by the secondary antibody conjugated HRP, which produces a light signal representative of the protein of interest. The membranes were incubated in either of the substrate solutions as necessary for 1 or 5 minutes (respectively). Finally, in a dark room and using a light proof AGFA Curix 60 film developer, chemiluminescence was detected using the CL-Xposure Film for various exposure times, as required. Alternately, chemiluminescence was detected using the syngene GeneGenome XRQ system.

2.2.10 Luciferase and β-galactosidase assays

Genetic reporter assays are frequently employed in molecular and cell biology for investigating various intracellular signalling pathways and gene expression regulation (Alam and Cook, 1990). Those involve the generation of a chimeric

reporter gene by cloning a regulatory region of a gene of interest upstream of the coding region of the reporter gene. Hence, the activity of the reporter would be a direct marker for that of the gene of interest. One of the most commonly used reporters is the firefly (*Photinis pyralis*) luciferase. This assay is based on quantifying the chemiluminescence produced upon the oxidation of luciferin to oxyluciferin by luciferase according to the equation shown in figure 2.1.

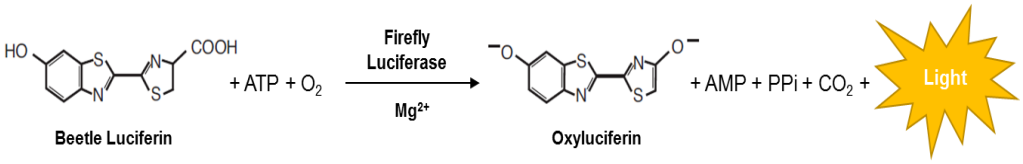


Figure 2.1: Firefly luciferase catalysis of Luciferin oxidation. Adapted from Promega, UK. https://www.promega.co.uk/-/media/files/resources/protocols/technical-bulletins/0/luciferase-assay-system-protocol.pdf?la=en.

2.2.10.1 Transfection

For the performance of luciferase and β-galactosidase assays, the MCF-7 ZEB1/ZEB2 GFP cell lines were firstly transfected with the appropriate luciferase reporter plasmids, β-galactosidase reporter plasmid and/or siRNAs (tables 2.7-8). The concentration of the luciferase reporters was 0.5-2μg/ transfection, β-galactosidase reporter was 0.7μg/transfection and the siRNAs were 100-150nM/ transfection. The transfections were performed as described previously (section 2.2.3)

2.2.10.2 Cells harvesting

Each plate was washed twice with PBS and the cells were then scraped with 1ml PBS and collected into a 1.5 microcentrifuge tube. After that, the cells were centrifuged at 1000rpm for 5min and the supernatants were discarded. The pellets were then subjected to three freeze and thaw cycles by placing the tubes on dry ice mixed with ethanol for 30sec then moved to a 37°C water bath for another 30sec. After that, the pellets were resuspended in 50μl of diluted 5x reporter lysis buffer (Promega, Southampton, UK) and incubated for 15min at room temperature. Then, the tubes were centrifuged again at full speed (13,200rpm) for 3 minutes and the supernatant was carefully removed, so not to disturb the pellet, and retained for luciferase and β-galactosidase activity analysis.

2.2.10.3 Luciferase assay

Initially, the luciferase substrate was prepared by resuspending the lyophilized substrate in luciferase assay buffer (Promega, Southampton, UK). Chemiluminescence was measured using a Sirius Single Tube Luminometer (Berthold Detection Systems GmbH, Germany) and following manufacturer instructions. Each sample supernatant was measured in triplicate, with 5µl of the sample being used for every measurement. For each measurement, 300µl of the luciferase assay solution was injected into each tube.

2.2.10.4 β-galactosidase assay

For each sample, 20µl of the supernatant was mixed with 150µl of β-galactosidase assay mastermix. The mastermix consists of the substrate ortho nitrophenyl β-galactoside (ONPG) (4mg/ml in 0.1M $NaPO_4$, pH7.5), $NaPO_4$ (0.1M and pH7.5) and Mg/β-ME (0.1M $MgCl_2$, 4.5M β-ME). The reaction was incubated at 37°C for 20-30min or until a faint brown colour develops. The absorbance was then measured at 405nm using a Sanyo Gallenkamp SPBIO Spectrophotometer.

2.2.10.5 Luciferase and β-galactosidase assay data analysis

To analyse the collected data, firstly, relative luminescence units (RLUs) for each sample was calculated by subtracting the background reading from the reading of each sample. After that, each sample was normalised against the corresponding β-galactosidase assay reading, which was performed to control the efficiency of the experiment. Then, fold change of activity was calculated by dividing the normalised values of the test samples by the control samples. Eventually, the test samples and the control samples were compared and analysed using unpaired two-tailed t-test. Statistical significance was approved for a difference of p-value of at least <0.05. All data analyses were done using Microsoft the software Excel (2010 version for windows) and/or GraphPad Prism (version 6.01). Data in the graphs are expressed as mean±SD. Asterisks in the figures designate statistical significance p-values where * refers to a p-value <0.05, ** refer to a p-value <0.01 and *** refer to a p-value <0.001.

2.2.11 Karyotyping of metaphase spreads

Karyotyping experiments were performed using a modified version of the protocol described in (Theunissen and Petrini, 2006). Cells were transfected (*siLIG1* or *siCDKN1B*) and/or treated (DOX) at the time of seeding and metaphase spreads preparation was performed at approximately 80% confluency. Cells were initially treated with colcemid at a final concentration of 0.5µg/ml for 1h or 2h (MCF-7 or A431 cells, respectively). Colcemid (aka demecolcine) is a drug related to the natural occurring compound colchicine, a derivative of the genus *Colchicum* plant, which interferes with microtubules assembly and spindle formation, resulting in blockage of mitosis at metaphase (Rieder and Palazzo, 1992). After that, cells were collected in 15ml tubes and washed with 2ml PBS which was also added to the tube. Then, to collect remaining attached cells, plate was trypsinised with 1ml of 2×TE which was deactivated with media from the 15ml tube.

Tubes were then centrifuged for 4 min at 1000rpm before removing most of the solution and resuspending the pellet in the remaining (~1ml). After that, 10ml of 0.075M of KCl, which must be made fresh on the day from 1M stock and pre-warmed to 37°C, was added to the tubes and left to swell at 37°C for 30min. From this stage forward, cells are very fragile, so, solutions were added very gently and dropwise against the wall of the tube with resuspending in around 2ml initially then adding the rest. After that, cells were gently resuspended in the KCl before adding 5ml of ice cold (3:1) methanol/acetic acid fixative, which was also made fresh on the day and stored at -20 beforehand. The tubes were then centrifuged for 4min at 1000rpm and the supernatant was aspirated off leaving around 1ml in which the cells were resuspended. After that, 10ml of fixative was added to the tubes followed by centrifugation again. This fixation step was repeated twice, to ensure the removal of any remaining salt, and the cells were placed overnight at -20°C.

In the following day, fixed cells were dropped on twin frosted microscope slides, so that the cells burst and release the chromosomes, and placed on an 80°C heat block covered with wet tissue papers. After that, slides were left to dry overnight before staining with freshly made 5% v/v Giemsa stain for 10min in a Coplin jar. Slides were then rinsed with water and placed on the heat block as

previous and left to dry overnight at RT. Finally, slides were mounted with the permount DPX and coverslips. Slides were then scanned to image metaphase spreads using Olymbus cyto-system and Case Data manager software version 6.0.

For each experiment, 30-50 spreads or approximately 2000 chromosomes were analysed. Spreads were analysed for chromosomal breaks, chromatid breaks and fusions. The percentage of each of the aberrations was then calculated in relation to the total number of chromosomes. Each experiment was performed in triplicate and then the test group was compared with the control group for statistical significance using unpaired student's t-test on GraphPad Prism (version 6.01). Data in the graphs are expressed as mean±SD. Asterisks in the figures designate statistical significance p-values where * refers to a p-value <0.05, ** refer to a p-value <0.01 and *** refer to a p-value <0.001.

Chapter 3 : ZEB proteins regulates DNA Ligase I expression through Rb-E2F

3.1 Introduction

ZEB2-induced EMT in the epidermoid carcinoma A431 cells resulted in the altered expression of multiple DNA repair genes including *LIG1*. Further work from our lab has proposed that this reduction in *LIG1* correlated with the reduced cyclin D1 expression in A431 cells undergoing an EMT (Dr Gina Tse, Dr Eugene Tulchinsky lab, university of Leicester, 2015). EMT reduced the expression of cyclin D1 through the direct repression of the *CCND1* promoter by ZEB2 (Mejlvang et al., 2007). The mechanism by which EMT induction and cyclin D1 repression caused LIG1 loss was still unknown.

Cyclin D1 can regulate *LIG1* transcription through two different mechanisms. The first mechanism would be through its canonical role in the regulation of the G1-S phase transition. This canonical role involves interaction with CDK4/6, initiation of Rb serial phosphorylation culminating in derepression of the E2F transcription factors, and the transcription of genes implicated in S-phase. The second mechanism involves cyclin D1 direct binding to DNA and regulation of gene transcription independently of cell cycle or E2Fs (Casimiro et al., 2012). The non-canonical function of cyclin D1 is in the maintenance of genomic stability, regulation of DNA damage repair (DDR) signalling and cell migration (Casimiro et al., 2012; Casimiro et al., 2014)

Data from this chapter shows that overexpression of either ZEB1/ZEB2 in the breast cancer cell line MCF-7 also results in LIG1 loss and reduction in cyclin D1 levels. This further suggests that LIG1 expression is dependent on cyclin D1. By using LIG1 promoter reporters and CDK4/6 inhibitors, I demonstrated that LIG1 expression is regulated by G1/S checkpoint.

3.2 Aims

- To test whether expression of cyclin D1 and LIG1 correlate in novel cell models of EMT based on breast carcinoma cell line MCF7.
- To investigate mechanism underlying this correlation

3.3 Results

3.3.1 ZEB1/2 induction reduces Cyclin D1 and DNA Ligase I (LIG1) expression in MCF-7 cells

To test if expression of LIG1 correlate with cyclin D1, I used two novel MCF-7 breast cancer cell lines with DOX regulated expression of ZEB1 or ZEB2. These cells were treated with DOX (2µg/ml) or left untreated for 72h, and then lysates were obtained. Immunostaining for LIG1 and cyclin D1 has validated the downregulation of the two proteins upon ZEB2 induction in two different clones of MCF7 cells (Figure 3.1). Moreover, inducing the expression of ZEB1 also caused downregulation of LIG1 and cyclin D1 expression levels (Figure 3.1). So, these data and previous studies indicate that repression of LIG1 and cyclin D1 by different EMT-TFs commonly occurs in different epithelial cancer cell lines.

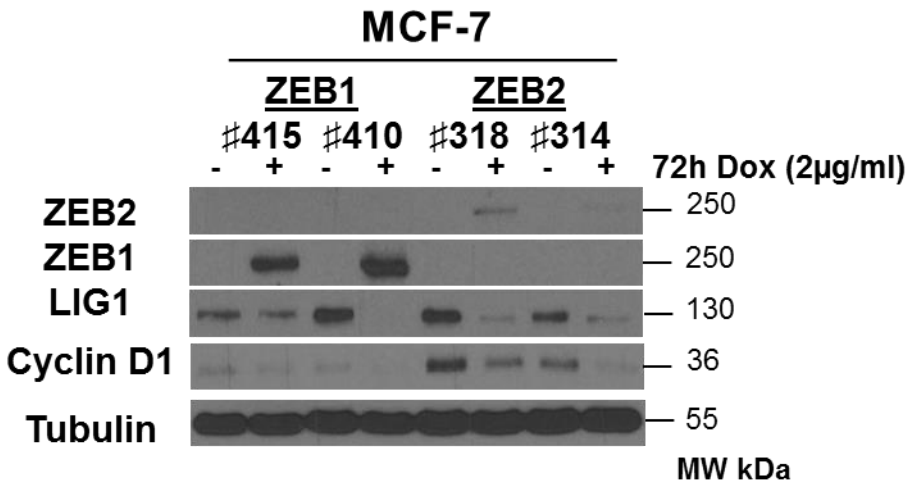


Figure 3.1: ZEB1 or ZEB2 induction results in the downregulation of LIG1 and cyclin D1.

Cells were cultured for 72h with/without Dox before lysing and processing for western blotting. It was observed that whenever cyclin D1 was reduced or absent, LIG1 was found to be reduced or absent too, suggesting a potential correlation. Tubulin was used as loading control. #X refers to the clone number. The PageRuler Plus™ Prestained protein ladder was used as a molecular weight (MW) marker. The experiment was repeated 3 times and a representative blot is shown.

3.3.2 Investigating how cyclin D1 regulates the LIG1 expression

Previous work has shown that ZEB2 regulates LIG1 expression primarily at the level of transcription (Dr Gina Tse, Dr Eugene Tulchinsky lab, the University of Leicester). Bioinformatics' analysis of the *LIG1* gene (UCSC Genome Browser) revealed two active chromatin regions in the promoter and first intron (a potential enhancer element) of the gene (Figure 1.13). To investigate if the activity of these elements is cyclin D1-dependent, three recombinant reporter plasmids were generated and used (Figure 3.2). In the first one, the *LIG1* promoter (1727bp) was subcloned into the pGL3-basic plasmid (4818bp) (Appendix 1). In the second construct, the intronic enhancer region (2532bp) of *LIG1* was inserted into the pGL3-SV40 promoter containing plasmid (5010bp) (Appendix 1). The last reporter was constructed by substituting the SV40 promoter in the pGL3-promoter LIG1 enhancer plasmid with the *LIG1* promoter. The first plasmid was provided by the lab whereas the other two plasmids were constructed for this project.

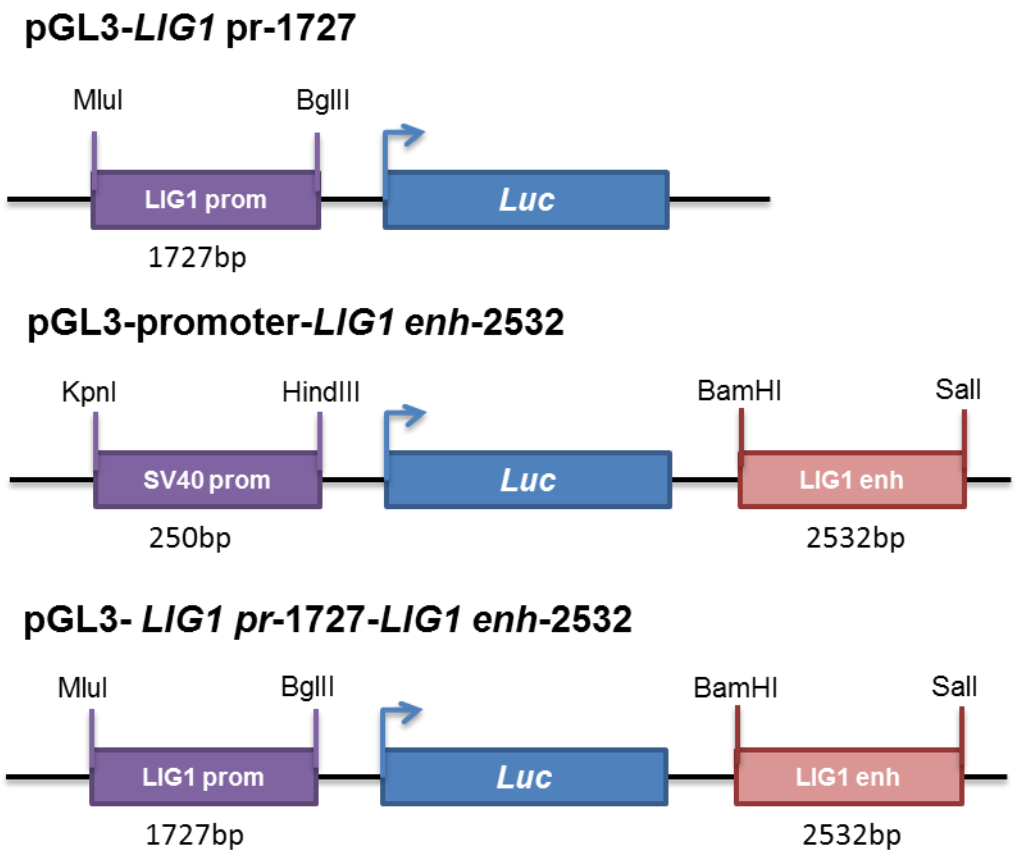


Figure 3.2: Schematic representation of the cloned plasmids of the *LIG1* gene.

Three recombinant reporter plasmids were generated to investigate the mechanism by which cyclin D1 and ZEB proteins regulate LIG1 expression. In the first one, the LIG1 promoter (1727bp) was subcloned into the pGL3-basic plasmid. In the second construct, the intronic enhancer region (2532bp) of LIG1 was inserted into the pGL3-SV40 promoter containing plasmid (5010bp). The last reporter was constructed by substituting the SV40 promoter in the pGL3-promoter LIG1 enhancer plasmid with the LIG1 promoter.

3.3.2.1 Generation of the pGL3-promoter-*LIG1 enh*-2532 and the pGL3- *LIG1 pr*-1727-*LIG1 enh*-2532

pGL3-promoter-*LIG1 enh*-2532

The enhancer region of the *LIG1* gene was amplified from the whole human genome and inserted into the vector between BamHI and SalI restriction sites. The lengths of the fragment in new constricts were as predicted, as confirmed by BamHI and SalI (Figure 3.3). The sequences were verified by GATC Biotech.

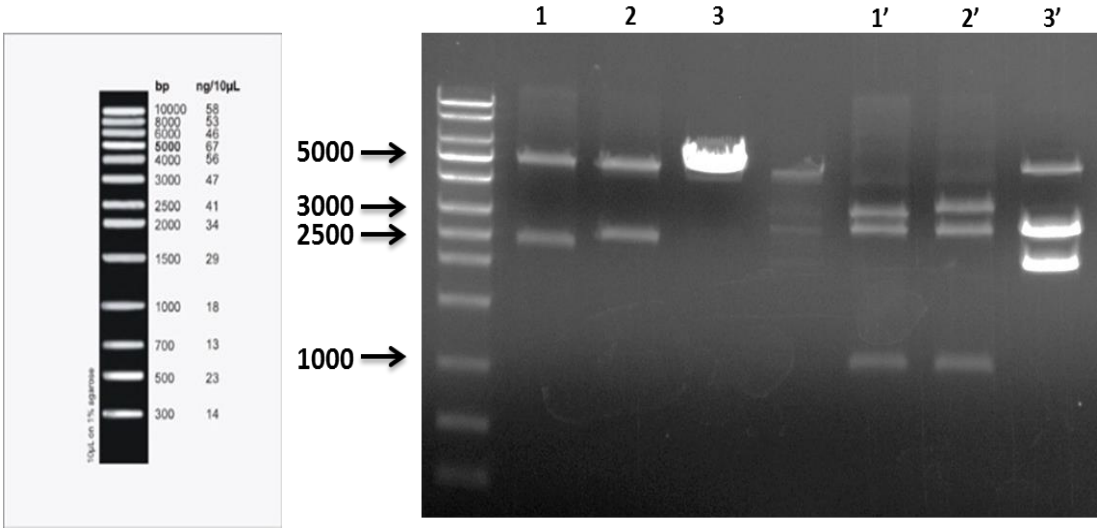


Figure 3.3: The pGL3-promoter-*LIG1 enh*-2532 recombinant plasmid confirmation.

Two samples of plasmid DNA were obtained from the transformation and were processed to confirm the success of the cloning (lanes 1, 2, 1' and 2') alongside the empty pGL3-promoter vector as a control (3 and 3'). 1, 2 and 3 were cleaved with BamHI and SalI, whereas 1', 2' and 3' were cleaved with SalI and BglII. Reactions were then analysed in 1% w/v agarose/TAE gel. The HighRanger 1 kb DNA Ladder was used as a marker.

pGL3- *LIG1 pr*-1727-*LIG1 enh*-2532

The promoter region of the *LIG1* was amplified from the pGL3-LIG1 pr-1727 plasmid and inserted into the vector between KpnI and HindIII restriction sites. The lengths of the fragment in new constricts were as predicted, as confirmed by BamHI and SalI as well as KpnI and HindIII (Figure 3.4). The sequences were verified by GATC Biotech.

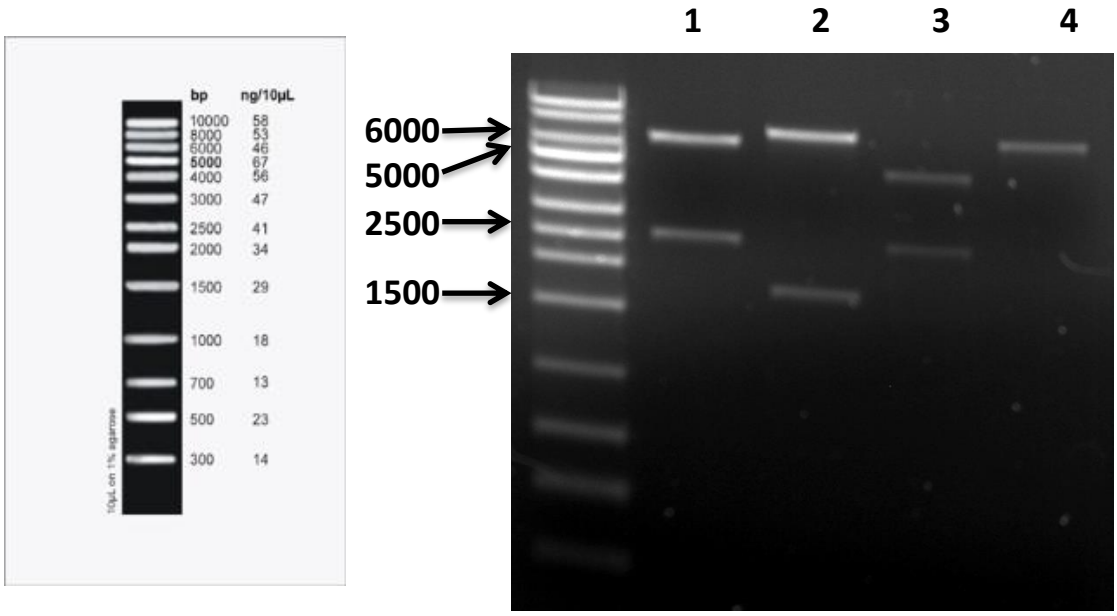


Figure 3.4: The pGL3- LIG1 pr-1727-LIG1 enh-2532 recombinant plasmid confirmation.

One clone obtained from the transformation was analysed to confirm the success of the cloning (lanes 1 and 2) alongside the pGL3-promoter-LIG1 enh-2532 vector as a control (3 and 4). 1 and 3 were digested with BamHI and SalI whereas 2 and 4 were digested with KpnI and HindIII. Reactions were then analysed in 1% w/v agarose/TAE gel. The HighRanger 1 kb DNA Ladder was used as a marker.

3.3.3 ZEB1/2 overexpression down-regulates *LIG1* promoter activity

To investigate the effect of ZEB1/2 overexpression on the activity of the *LIG1* promoter, the MCF-7/ZEB1 or MCF-7/ZEB2 cells were transfected with pGL3-*LIG1 pr*-1727 and cultured in the presence or absence of DOX. ZEB1/2 overexpression resulted in approximately 2-fold reduction in *LIG1* promoter activity in MCF-7/ZEB1 and around 1.6-fold reduction in MCF-7/ZEB2 cells (p-values <0.01 and <0.001, respectively) (Figure 3.5, A and B). Then, the MCF-7 ZEB1 cells were transfected with either the pGL3- *LIG1 pr*-1727-*LIG1 enh*-2532 plasmid or the pGL3-*LIG1 pr*-1727 to determine how ZEB1/2 overexpression would affect the promoter activity in each case. Presence of the putative enhancer element in the construct did not enhance the effect of ZEB1 on the reporter activity (Figure 3.6). This indicated that the *LIG1* promoter, rather than intronic sequences were responsible for the *LIG1* regulation by ZEB1/2 overexpression. Therefore, pGL3- LIG1 pr-1727 construct was used in the next experiments.

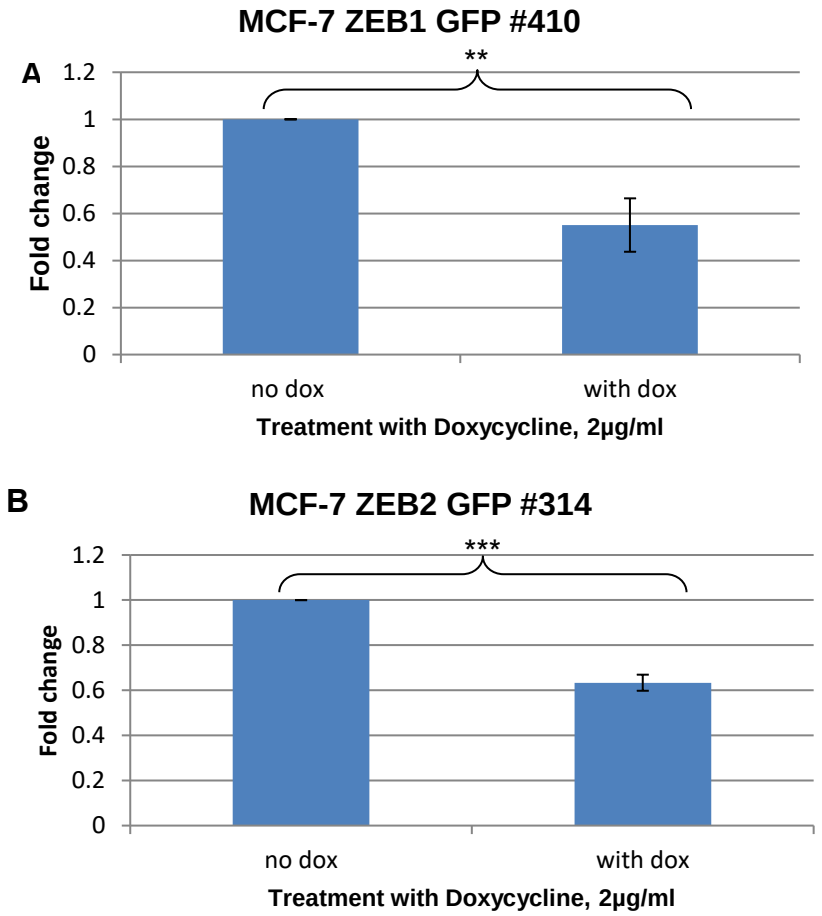


Figure 3.5: Effect of ZEB1/2 overexpression on the sensitivity of the *LIG1* promoter in MCF-7 cell lines.

The graphs show the decrease in *LIG1* promoter activity in MCF-7 ZEB1 GFP cell line (A) and MCF-7 ZEB2 GFP cell line (B) transfected with 2µg of pGL3-*LIG1 pr*-1727 and maintained for 72h in media with/without 2µg/ml Doxycycline. Cells were also transfected with a β-galactosidase reporter plasmid that is used for luciferase activity normalisation. Asterisks refers to statistical significance ** refers to p-value <0.01 (0.0024) and *** refers to p-value <0.001 (0.0001). Data are plotted as mean±SD. #X refers to the clone number. N=3.

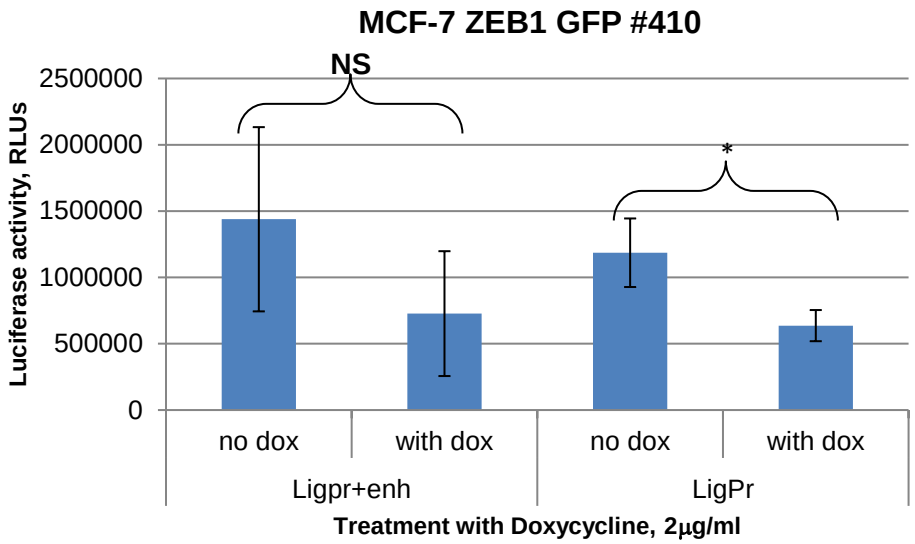


Figure 3.6: *LIG1* gene promoter, but not a putative intronic enhancer confers reporter responsiveness to ZEB1 overexpression.

The graph compares the effect of ZEB1 overexpression on the activity of *LIG1* promoter alone (pGL3-*LIG1* pr-1727) or coupled to the putative enhancer (pGL3- *LIG1* pr-1727-*LIG1* enh-2532). Cells were transfected with one of the plasmids along with a β-galactosidase reporter plasmid and maintained in media with/without 2µg/ml Dox for 72h. The presence of the putative enhancer element in the construct did not enhance the effect of ZEB1 on the reporter activity, suggesting that the promoter is the responsive element. **NS** non-significant. * p-value <0.05 (0.03). Data are plotted as mean±SD. #X refers to the clone number. N=3.

3.3.4 Effect of cyclin D1 knockdown on the activity of the *LIG1* promoter

LIG1 is active in S phase of the cell cycle, where it plays a role in DNA replication and repair through HR pathway. As canonical cyclin D1 function is in the control of G1-S transition, and expression of cyclin D1 is down-regulated by ZEB1, we aimed to test whether *LIG1* promoter activity is dependent on cyclin D1 presence in the cells. MCF-7/ZEB1 or MCF-7/ZEB2 cells were transfected with small interfering RNA (siRNA) targeting cyclin D1 (si*CCND1*) alongside with the pGL3-*LIG1 pr*-1727. Cyclin D1 knockdown resulted in approximately 2.2-fold reduction in the *LIG1* promoter activity in the MCF-7/ZEB1 cells compared to ~7-fold reduction in the MCF-7/ZEB2 cells. Both reductions were statistically significant (p-values <0.0001 and <0.05, respectively) (Figure 3.7, A and B). Western blotting was used to confirm cyclin D1 knockdown which showed a complete loss of cyclin D1 as well as LIG1 expression in both cell lines (Figure 3.7, A and B). To confirm that the cyclin D1 knockdown effect was specific to the *LIG1* promoter, si*CCND1* was also transfected along with the SV40 promoter (pGL3-promoter plasmid). There was no statistically significant difference between the SV40 promoter activity when transfected with the siCONTROL or siCCND1 (p-value > 0.05) (Figure 3.8).

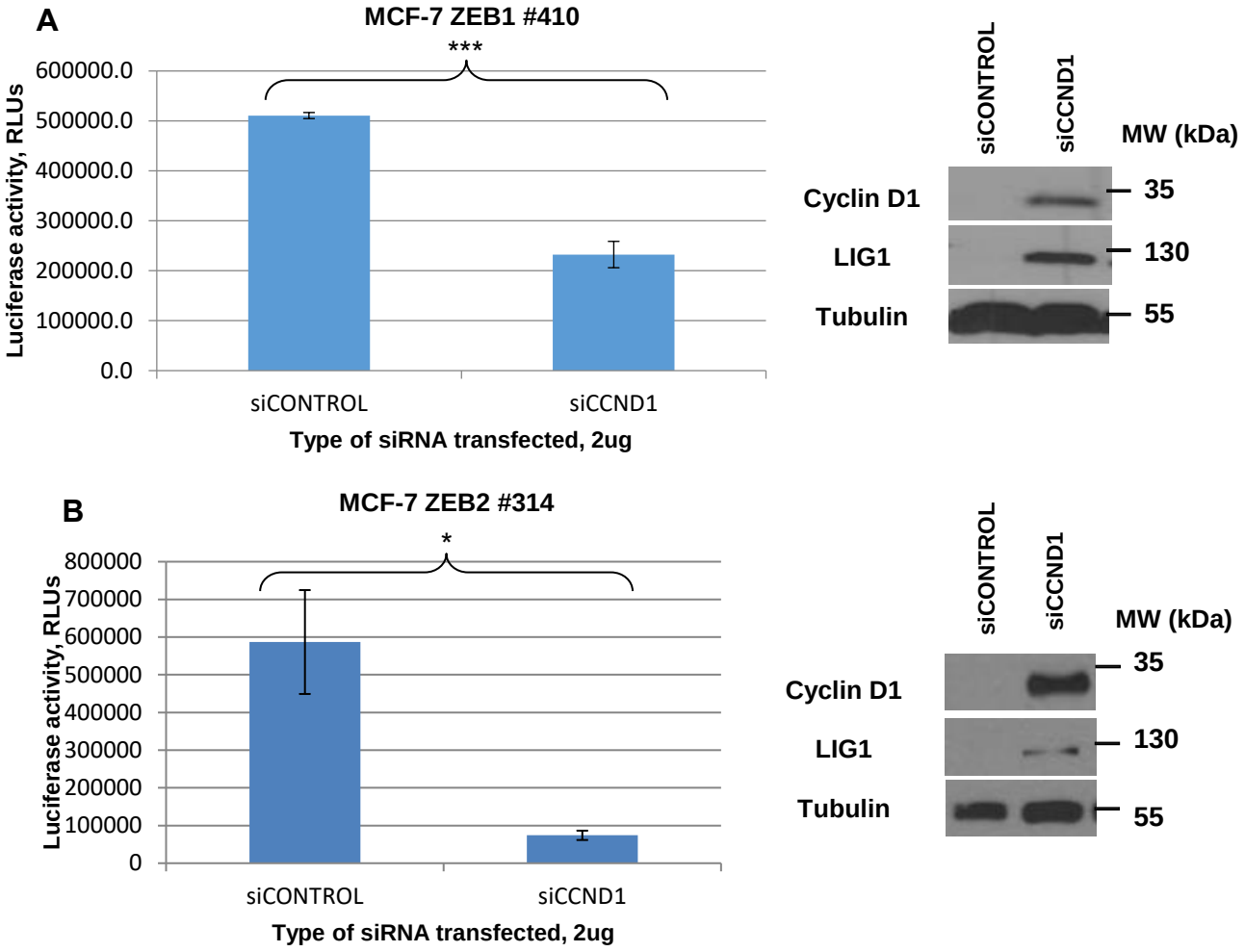


Figure 3.7: Effect of cyclin D1 knockdown on the activity of the *LIG1* promoter in MCF-7 cell lines.

The graphs show the reduction in the activity of the *LIG1* promoter in MCF-7/ZEB1 cell line (A) and MCF-7/ZEB2 cell line (B) transfected with 2µg of either siCCND1 or siCONTROL along with 1µg pGL3-*LIG1 pr*-1727 plasmid and a β-galactosidase reporter plasmid. These cells were maintained for 72h before lysis and the analysis of luciferase activity. β-galactosidase assay was used for luciferase activity normalisation. Cyclin D1 down-regulation was confirmed by western blots for the two cell lines (A) and (B), respectively. Western blots also show loss of LIG1 in cells with depleted cyclin D1. * p-value <0.05 (0.0030). *** p-value <0.001 (0.0001). Data are plotted as mean±SD. #X refers to the clone number. N=3.

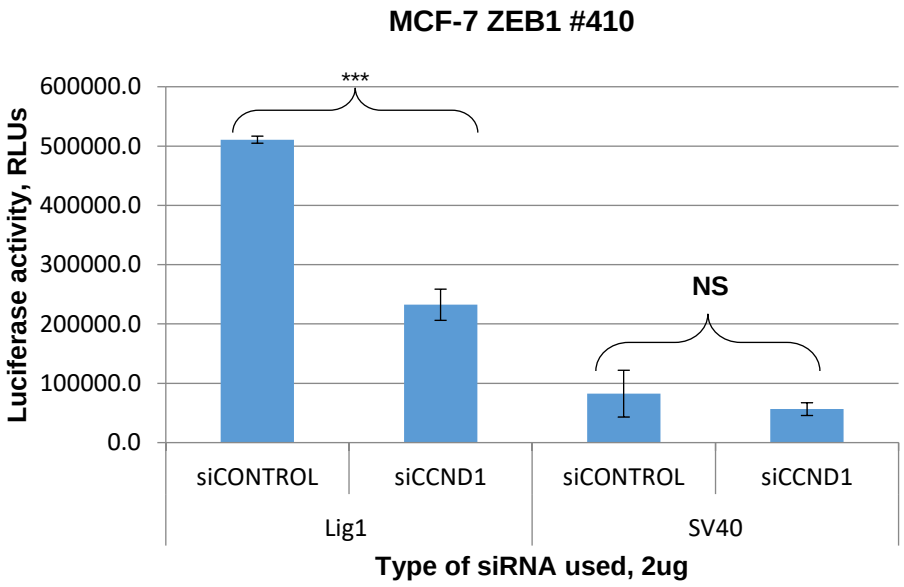


Figure 3.8: Cyclin D1 knockdown has no effect on SV40 promoter in MCF-7/ZEB1 cells.

The graph compares the activity of *LIG1* and SV40 promoters in the MCF-7/ZEB1 cell line transfected with either si*CCND1* or siCONTROL along with 1µg of either pGL3-*LIG1 pr*-1727 plasmid or pGL3-promoter plasmid and a β-galactosidase reporter plasmid. These cells were maintained for 72h before lysis and the analysis of luciferase activity. β-galactosidase assay was used for luciferase activity normalisation. *** p-value <0.001 (0.0001). **NS**, non-significant. Data are plotted as mean±SD. #X refers to the clone number. N=3.

3.3.5 Shorter fragment of *LIG1* promoter DNA confers ZEB1 responsiveness

We analysed the *LIG1* promoter sequence using the web-based tool ALGGEN-PROMO for potential transcription factor binding sites. The *LIG1* promoter was found to contain around 461 potential transcription factors binding sites, with around 68 binding sites for members of the E2F family of transcription factors. Interestingly, 43 of these binding sites were for the transcription factor E2F-1, the transcription factor that is activated because of Rb phosphorylation by the CDK4/6-Cyclin D and CDK2-Cyclin E complexes (see section 1.4.4.2.1). An enrichment of these binding sites was specifically observed in the proximal 500bp of the *LIG1* promoter sequence (Figure 3.9). For this reason, the 500bp of the promoter DNA was amplified and cloned into the pGL3-basic vector upstream the luciferase reporter gene (Figure 3.10).

Figure 3.9: The putative binding sites for E2F-1 in the promotor region of *LIG1*.

The sequence of LIG1 promotor was uploaded to ALGGEN-PROMO to determine putative transcription binding sites. Many of the predicted binding sites were for E2F-1, which are highlighted in the figure. The binding sites were more frequent in the proximal region of the *LIG1* promotor. Figure is from (Dr Gina Tse, Dr Eugene Tulchinsky Lab).

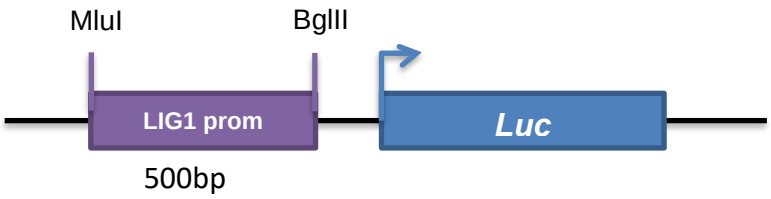


Figure 3.10: Schematic representation of the cloned pGL3-*LIG1 pr*- 500 plasmid.

The proximal 500bp of the LIG1 promoter was found to be enriched for E2F-1 binding sites. So, it was cloned into the pGL3-basic vector (Appendix 1, A) to map the sensitivity of the promoter to ZEB1.

3.3.5.1 Generation of the pGL3-*LIG1 pr*- 500 plasmid

The 500bp region of the *LIG1* promoter was amplified using the pGL3-*LIG1 pr*-1727 plasmid DNA as a template and inserted into the vector between MluI and BglII restriction sites. The lengths of the fragment in new constricts were as predicted, as confirmed by MluI and BglII (figure 3.11). The sequence was verified by GATC Biotech.

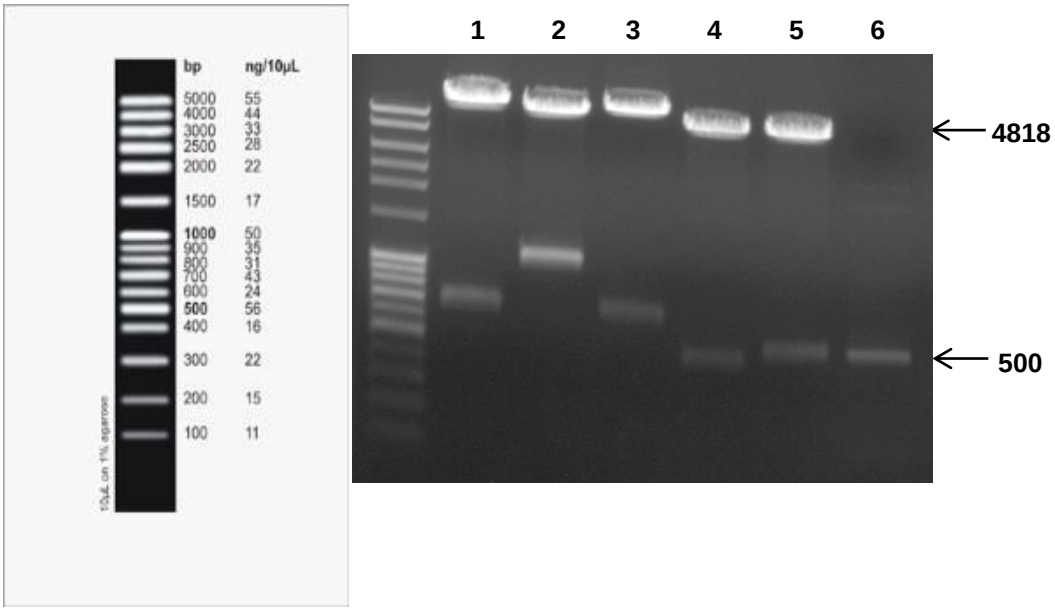


Figure 3.11: The pGL3-*LIG1 pr*- 500 plasmid confirmation.

Two clones were analysed to confirm the success of the cloning (lanes 4 and 5). Amplified 500bp fragment of the LIG1 promoter DNA was loaded in lane 6 as a control. 4-6 samples were digested with MluI and BglII. Lanes 1-3 were to confirm another plasmid. Reactions were then analysed in 1% w/v agarose/TAE gel. The FullRanger 100bp DNA Ladder was used as a marker.

3.3.6 Proximal 500bp fragment of the *LIG1* promoter confers ZEB responsiveness to a luciferase reporter in MCF7 cells

To determine the effect of ZEB proteins overexpression on the activity of the proximal 500bp DNA fragment of the *LIG1* promoter, MCF-7/ZEB1 or MCF-7/ZEB2 cells were transfected with pGL3-*LIG1 pr*- 500 or pGL3-LIG1 pr-1727 plasmid DNAs and cultured for 72h in the presence or absence of DOX. 500bp fragment exhibited the same level of responsiveness to ZEB1 and ZEB2 proteins as the larger 1727bp *LIG1* promoter fragment. ZEB1 and ZEB2 reduced reporter activities 2- and 1.6-fold respectively. Effects were statistically significant, p-values <0.001and <0.01, respectively) (Figure 3.12, A and B).

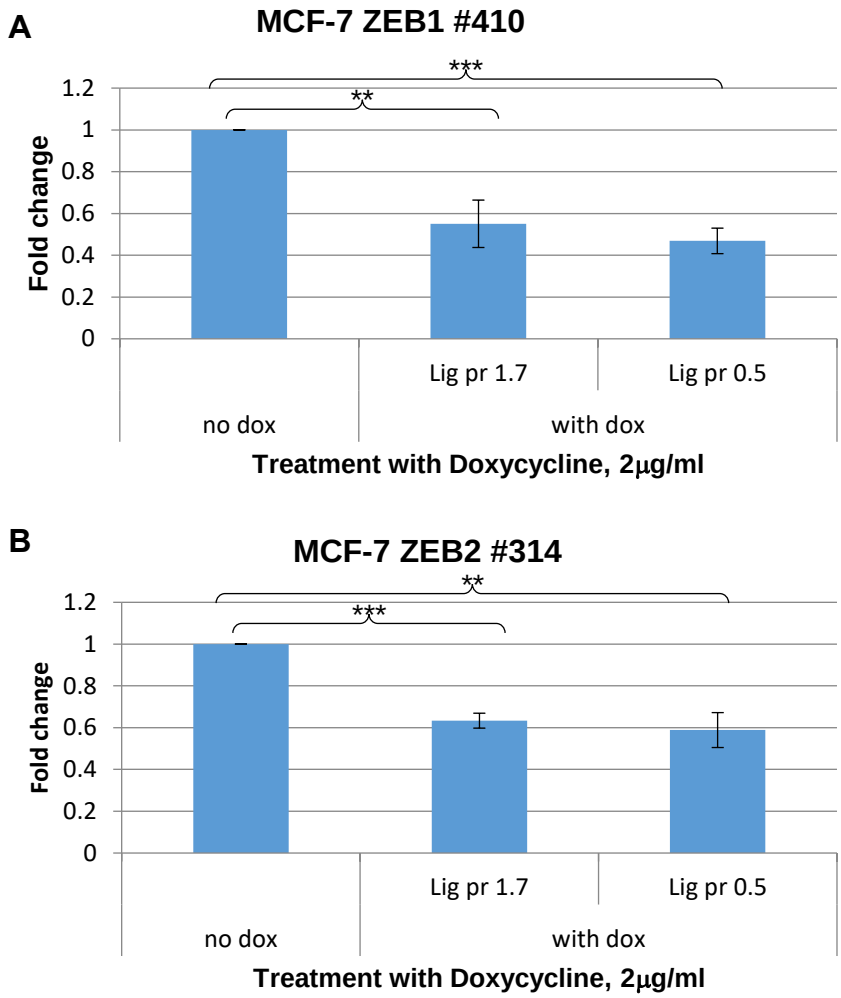


Figure 3.12: comparison of the effects of EMT on the sensitivity of the *LIG1* promoter fragments (500bp and 1727bp) in MCF-7 cell lines.

Cells were transfected with either pGL3-*LIG1 pr*- 500 or pGL3-*LIG1 pr*-1727 along with a β-galactosidase reporter plasmid and cultured for 72h in the presence or absence of DOX, to determine the effect of ZEB proteins overexpression on the activity of the proximal 500bp DNA fragment of the *LIG1* promoter. The graphs show that the 500bp fragment exhibited the same level of responsiveness to ZEB1 and ZEB2 proteins as the larger 1727bp *LIG1* promoter fragment. β-galactosidase assay was used for luciferase activity normalisation. ** p <0.01 (ZEB1=0.0024, ZEB2=0.001) and ***p<0.001 (ZEB1=0.0001, ZEB2<0.0001). Data are plotted as mean±SD. #X refers to the clone number. N=3.

3.3.7 Effect of cdk4/6 IV (CINK4) inhibitor on LIG1 expression

Expression level of LIG1 was strongly reduced by cyclin D1 knockdown. In addition, *LIG1* promoter was enriched for E2F-1 binding site, and its activity was dependent on the presence of cyclin D1. Therefore, we hypothesised that G1 checkpoint regulates *LIG1* expression. To test this hypothesis, MCF-7/ ZEB1/ZEB2 cells were treated with the cdk4/6 IV inhibitor (CINK4). CINK4 inhibitor is a triaminopyrimidine compound that acts as a reversible and ATP-competitive inhibitor of cyclin D1-bound cdk4/6 (Soni et al., 2001). Treatment with the CINK4 inhibitor resulted in the reduction in the LIG1 expression concomitant with Rb hypophosphorylation in both cell lines (Figure 3.13). The extent of the reduction was proportional to the concentration of the inhibitor, however, concentrations >4µM resulted in cytotoxicity. This implies a role for RB/E2F-1 in the control of LIG1 expression induced by ZEB proteins overexpression.

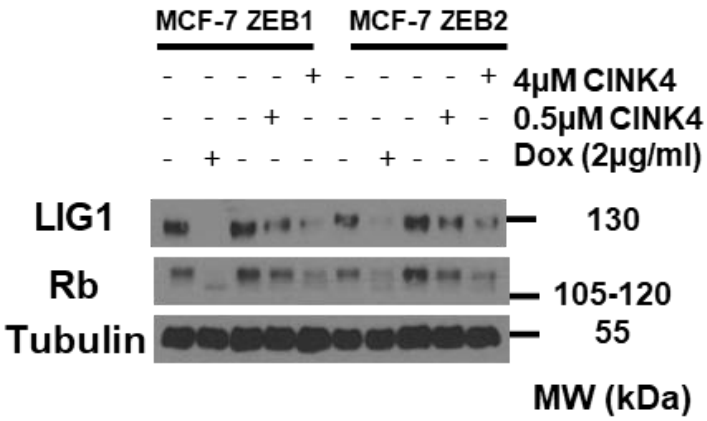


Figure 3.13: Effect of CINK4 inhibitor on the expression of *LIG1* in MCF-7 cells.

Cells were seeded with/without Dox for 72h with the addition of the cdk4/6 IV inhibitor (CINK4) to Dox untreated cells after 24h of seeding. At 72h, cells were lysed for western blotting. Cells were then stained for LIG1 and Rb to test whether LIG1 is G1-checkpoint controlled. Rb hypophosphorylation (lower band) was found to be concomitant with DNA Ligase I reduction. Upper band corresponds to hyperphosphorylated Rb (pRb). Tubulin was a loading control. The experiment was repeated 3 times and a representative blot is shown.

3.4 Discussion

Expression of LIG1 was downregulated by ZEB2 in the epidermoid carcinoma A431-ZEB2 and the colorectal cancer DLD-1-ZEB2 cell lines (Dr Gina Tse, Dr. Eugene Tulchinsky Laboratory). This downregulation was suggested to correlate with the reduction in cyclin D1 expression levels that resulted from direct interaction between ZEB2 and *CCND1* promoter (Mejlvang et al., 2007). This reduction of the LIG1 by ZEB2 correlates with the accumulation of ssbs observed previously in A431 treated with low dose of UV light (Dr Hussain Abbas, The University of Leicester). If not repaired, these ssbs would be converted to dsbs and contribute in genomic instability. However, the mechanism by which ZEB2 (i.e. EMT) regulates LIG1 expression has not yet been unravelled. So, here we aimed to examine mechanism of this regulation but with first studying other EMT models. We have validated the same correlation between the expressions of the LIG1 and cyclin D1 in two novel cell lines MCF-7/ZEB1 and MCF-7/ ZEB2, which express ZEB proteins in Dox-inducible manner.

Bioinformatics' search of *LIG1* revealed two active chromatin regions in gene promoter and in the first intron (a putative enhancer) (Figure 1.13). The promoter and the putative enhancer regions of the *LIG1* were cloned into the luciferase reporter plasmids pGL3-basic and pGL3-promoter, respectively, to analyse their activity. Luciferase assay experiments have shown that the *LIG1* promoter (i.e. pGL3-*LIG1 pr*-1727) is active and affected by ZEB proteins overexpression (i.e. EMT) in both of MCF-7/ZEB1 and MCF-7/ZEB2 cell lines (Figure 3.7 A&B). ZEB proteins overexpression by Dox has resulted in 1.6-2 fold reduction in *LIG1* promoter activity in the two cell lines. However, no significant difference has been observed when the intronic sequence was introduced in the enhancer region (i.e. pGL3-promoter-*LIG1* enh-2532 and pGL3-*LIG1* pr-1727-*LIG1* enh-2532) (Figure 3.8). So, from this point the promoter region was the focus of our study.

Next, cyclin D1 knockdown by RNAi has confirmed that the reduction in the *LIG1* promoter activity was due to the reduction in cyclin D1 expression. Knocking down cyclin D1 has resulted in ~2.2 and ~7 fold decrease in the *LIG1* promoter activity in the MCF-7 ZEB1/ZEB2 cells, respectively (Figure 3.9 A&B). This effect was not observed when cells were transfected with the control luciferase reporter

plasmid carrying SV40 promoter (figure 3.10). This implies that the effect of cyclin D1 knockdown is *LIG1* promoter-specific.

LIG1 promoter sequence analysis for potential transcription factor binding sites has shown an enrichment for E2F-1 binding sites (TTTc/gGCGCc/g) in the proximal 500bp fragment of the sequence (Figure 3.11). E2F-1 is the transcription factor activated as a result of Rb phosphorylation by the CDK4/6-Cyclin D and CDK2-Cyclin E complexes at G1-S transition and is implicated in the transcription of the genes essential for DNA replication during the S-phase (see section 1.4.4.2.1). For that reason, the 500bp fragment of the promoter was cloned into the pGL3-basic luciferase reporter plasmid. The responsiveness of this part of the promoter was very similar to that of the complete (1727bp) promoter. Indeed, ZEB proteins overexpression and EMT induction resulted in ~2 and ~1.6-fold reduction in the reporter activity in MCF-7 cell lines transfected with the pGL3-*LIG1* pr-500 plasmid (Figure 3.14 A&B). So, this suggests that the responsiveness of the *LIG1* promoter lies within the 500bp proximal promoter fragment.

To investigate whether the effect of cyclin D1 on the *LIG1* promoter is exerted indirectly via Rb and E2F or directly (e.g. binding of cyclin D1 to the promoter DNA), MCF-7 ZEB1/ZEB2 cells were incubated with the cdk4/6 IV (CINK4). CINK4 inhibitor is a triaminopyrimidine compound that acts as a reversible and ATP-competitive inhibitor of cdk4/6 (Soni et al., 2001). As a result, it interferes with the hyperphosphorylation of Rb (hence, G1-S transition and E2F release). The expression of LIG1 was found to be downregulated whenever Rb appeared hypophosphorylated in both cell lines treated with CINK4 (Figure 3.15). This suggests the dependency of *LIG1* expression on E2F. This is consistent with the recent study that suggests the cell cycle regulation of genes involved in ssbs. The study has shown that genes involved in early and late steps of BER and those of late NER, including *LIG1*, are upregulated either in G1/S phase or S phase, respectively (Mjelle et al., 2015). So, here we suggest that E2F induces the expression of *LIG1* and probably through direct interaction with its promoter. Further experiments, such as chromatin immunoprecipitation should be carried out to confirm this interaction.

Although ssbs mainly occurs in the G1 phase, the long life of LIG1 protein as demonstrated previously (Dr Gina Tse, Dr. Eugene Tulchinsky Laboratory, the University of Leicester) would enable the cycling cell to accumulate enough protein for the repair of ssbs. In fact, cells must be withdrawn from the cell cycle for a long time (72h) to lose the LIG1 expression resulting in accumulation of ssbs. Activation of LIG1 at the G1/S checkpoint is consistent with the functions of LIG1 in DNA replication (ligating the Okazaki fragments of the lagging strand) (Singh et al., 2014). Mouse embryonic stem cells with deletion of *LIG1* or other ligases have been reported to be inviable (Frank et al., 2000; Petrini et al., 1995; Puebla-Osorio et al., 2006). However, this has recently been debated in the reports suggesting compensation for LIG1 by LIG3 (Arakawa et al., 2012; Han et al., 2014). Experiments in chicken DT40 cells have demonstrated that LIG3 and LIG1 function with the same efficiency in NER (Paul-Konietzko et al., 2015). Additionally, other recent studies performed in mouse B lymphocyte (CH12F3) and chicken DT40 Bursal lymphoma cell lines with deleted *LIG1* were viable (Arakawa et al., 2012; Han et al., 2014). *LIG1*-null CH12F3 cells exhibited no defects in proliferation, genotoxic drugs sensitivity, or chromosomal integrity (Han et al., 2014). Therefore, LIG1 and LIG3 are functionally redundant in DNA replication in chicken DT40 (Arakawa et al., 2012). All these studies have used somatic B cell lines (i.e. chicken DT40 and CH12F3 cells. This could suggest that functional redundancy between DNA ligases exists in somatic cells, but not in embryos where all the DNA ligases are vital.

As indicated above, LIG1 is involved in long-patch BER, NER, HR and DNA replication. Therefore, reduction in its expression in the G1 phase may result in the accumulation of the ssbs that will be converted to dsbs in the S phase when DNA replication occurs. As the homologous recombination (HR) process is defective as well upon EMT induction (Dr Gina Tse, Dr. Eugene Tulchinsky Laboratory), those breaks will be repaired by the less accurate non-homologous end joining (NHEJ) leading to chromosomal instability and an increase in genomic instability.

Chapter 4 : ZEB1 induces cell cycle arrest through p27Kip1 upregulation

4.1 Introduction

EMT-TFs induce cell cycle arrest at either G1 or G2 phases. The G1 phase cell cycle arrest could occur as a result of either the downregulation of cell cycle positive regulators (CDKs and cyclins), or the upregulation of the cell cycle negative regulators (e.g. CDK-Is p21Cip1 and p27Kip1). For instance, snail expression and EMT induction for 72h in MDCK cells resulted in cell cycle arrest in G1 phase. This was caused by upregulation of p21Cip1 as well as the downregulation of cyclin D2 expression through direct repression of the *CCND2* gene via E-box consensus regions (Vega et al., 2004). ZEB2 expression in the squamous carcinoma cell line A431 also resulted in cell cycle arrest in G1 phase. However, this was shown to occur through direct repression of the *CCND1* promoter (Mejlvang et al., 2007). Moreover, twist or snail expression in tubular epithelial cells (TECs) of the kidney induced a cell cycle arrest in G2 phase via the upregulation of p21Cip1. This is to impact on the ability of the cells to repair and regenerate, which leads to kidney fibrosis (Lovisa et al., 2015).

In this Chapter I demonstrated that ZEB1 overexpression in MCF-7 cells also resulted in cell cycle arrest that is caused by a strong upregulation of p27Kip1 expression. RNAi and growth curve experiments confirmed this link, as p27Kip1 knockdown rescued ZEB1-induced cell cycle arrest but without affecting mesenchymal features. However, p27Kip1 knockdown was not enough to reverse an effect of ZEB1 and EMT on LIG1, cyclin D1 and Rb. This suggests that p27Kip1 upregulation merely uncouples quiescence from other features of ZEB1-induced EMT. qPCR analysis and reporter assays suggest that ZEB1 regulates *CDKN1B* transcriptionally through the promoter region.

4.2 Aims

- Elucidate how ZEB overexpression (i.e. EMT) induces cell cycle arrest in MCF-7
- Examine whether the arrest in MCF-7 cells is linked to p27Kip1
- Examine how ZEB overexpression regulates p27Kip1 expression

4.3 Results

4.3.1 p27Kip1 upregulation causes cell cycle arrest in ZEB1 expressing MCF-7 cells

Cell cycle distribution analysis has shown that ZEB1 overexpression in MCF-7 cells resulted in cell cycle arrest at G1 and a very high reduction of G2. ZEB1 overexpression was found to increase the G1 peak (P3) from approximately %60 in uninduced cells to approximately %85 in Dox-induced cells (Figure 4.1, A & B). ZEB1 overexpression also resulted in the reduction of the G2 peak (P5) from around %26 in uninduced cells to around %11 in Dox-induced cells. Same with the S-phase peak (P4) that was reduced from around %14 to around %4 after Dox-induced ZEB1 overexpression. This was a reminiscent of what was shown in A431 cells after EMT induction by ZEB2. Cell cycle arrest at G1 in A431 cells is thought to be a result of a direct repression of cyclin D1 expression by ZEB2 (Mejlvang et al., 2007). Therefore, cyclin D1 expression was examined in MCF-7 cells to identify whether cell cycle arrest occurs through the same route.

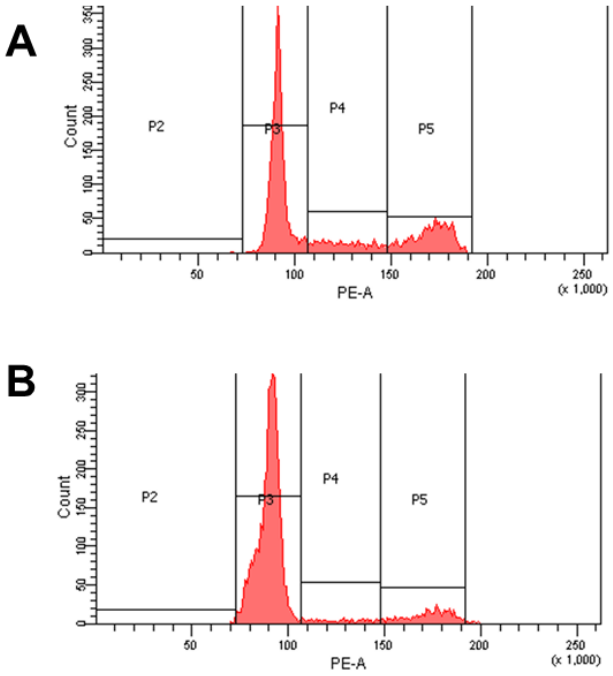


Figure 4.1: ZEB1 overexpression in MCF-7 cells resulted in cell cycle arrest at G1 phase.

FACS analysis of MCF-7/ZEB1 cells maintained in the absence (A) or presence (B) of DOX for 72h. P2: necrotic cells, P3: G1-phase, P4: S-phase, P5: G2-phase. This figure is a representative of triplicate experiments.

In addition to the effect on cyclin D1 expression (Figure 3.1), we found that ZEB1 overexpression strongly upregulated the CDK inhibitor p27Kip1 (Figure 4.2). This suggested a potential involvement of p27Kip1 in MCF-7 cell cycle arrest. In fact, EMT induction by ZEB2 in epidermoid carcinoma (A431-ZEB2) also resulted in p27Kip1 upregulation.

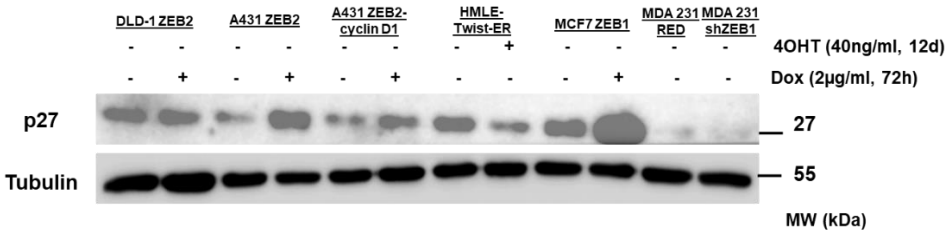


Figure 4.2: p27KIP1 tumour suppressor protein is overexpressed in the course of ZEB overexpression (i.e. EMT) in carcinoma cell lines but not HMLE-Twist-ER.

Cells were cultured with or without DOX for 72 h (120h for DLD-1 ZEB2 cells). HMLE cells were cultured with or without 4OHT (tamoxifen) for 12d before lysing. The expression of indicated proteins was analysed by Western blotting. Tubulin was used as a loading control. The experiment was repeated 3 times and a representative blot is shown.

4.3.2 Loss of p27KIP1 expression results in proliferating ZEB1-overexpressing cells

To examine whether p27Kip1 upregulation was implicated in ZEB1-induced cell cycle arrest in MCF-7 cells, p27Kip1 was knocked down by siRNA. Cells overexpressing ZEB1 with depleted p27Kip1 displayed a typically flat mesenchymal phenotype with protruding filopodia (Figure 4.3). Interestingly, proliferation rate of these cells was very similar to that of epithelial MCF-7 cells, as demonstrated by growth curve of these cells (Figure 4.4).

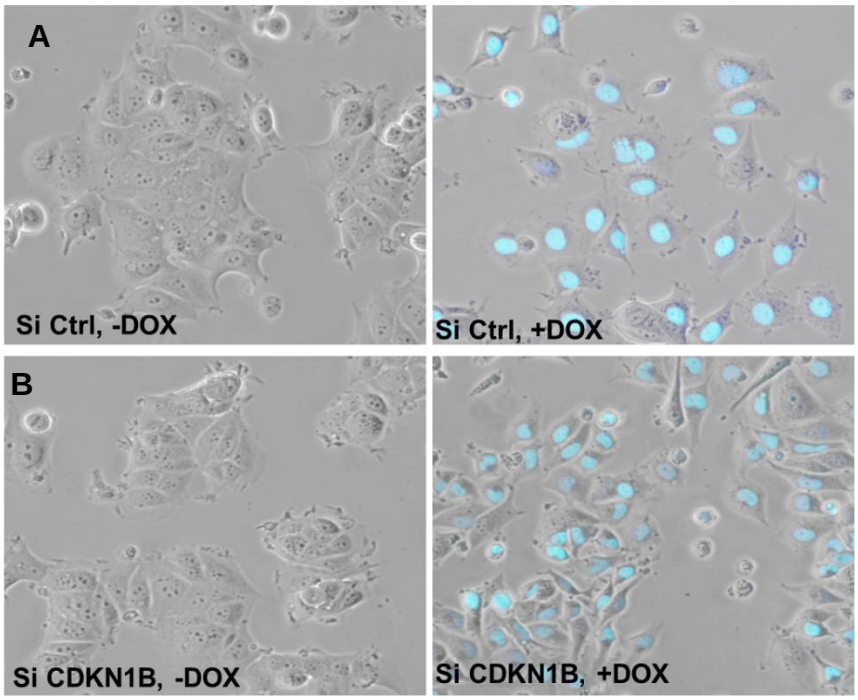


Figure 4.3: Loss of p27KIP1 expression by siRNA abolished the quiescence of ZEB1 overexpressing (i.e. EMT induced) cells (**B**, right panel) **as compared to control ZEB1 overexpressing cells** (**A**, right panel). Cells were transfected with either a control siRNA (siCtrl) or 150nmol of si*CDKN1B* and cultured with or without 2µg/ml Dox for 72h before imaging. Fluorescence and phase contrast images were generated and merged using ImageJ. The experiment was repeated twice with similar results.

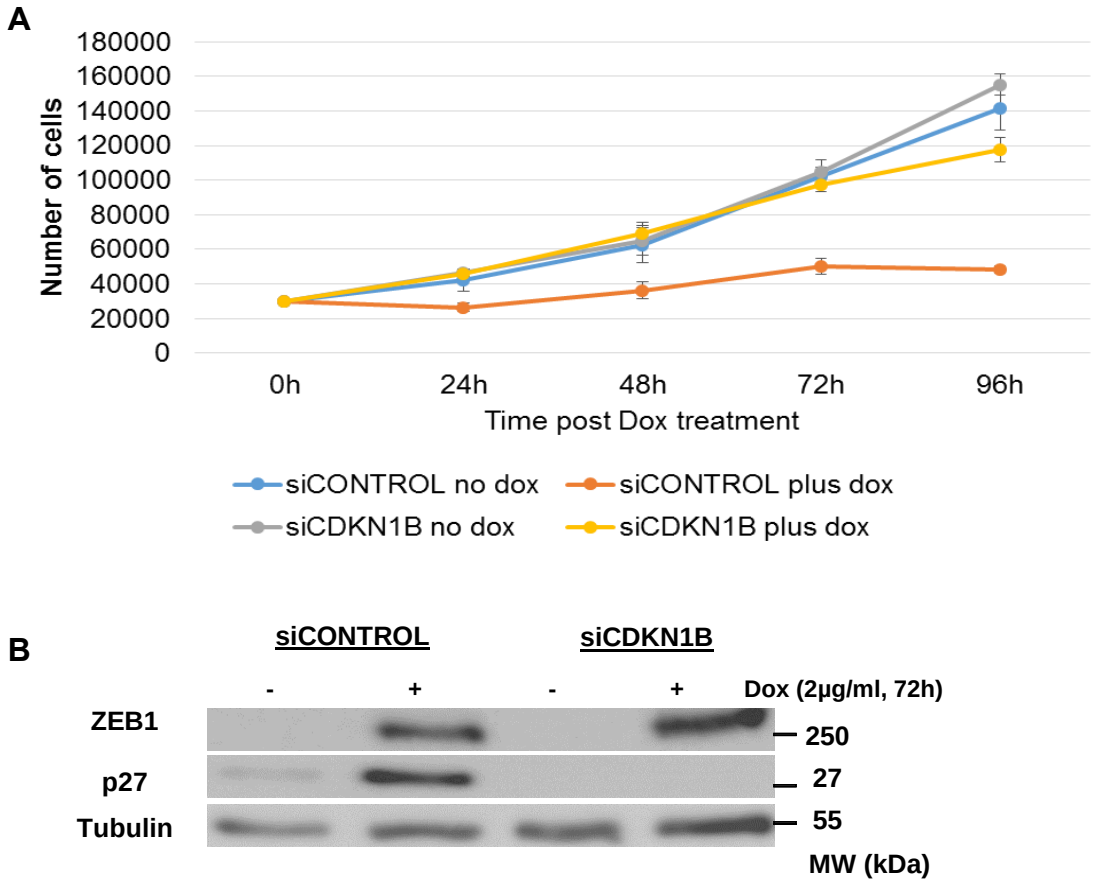
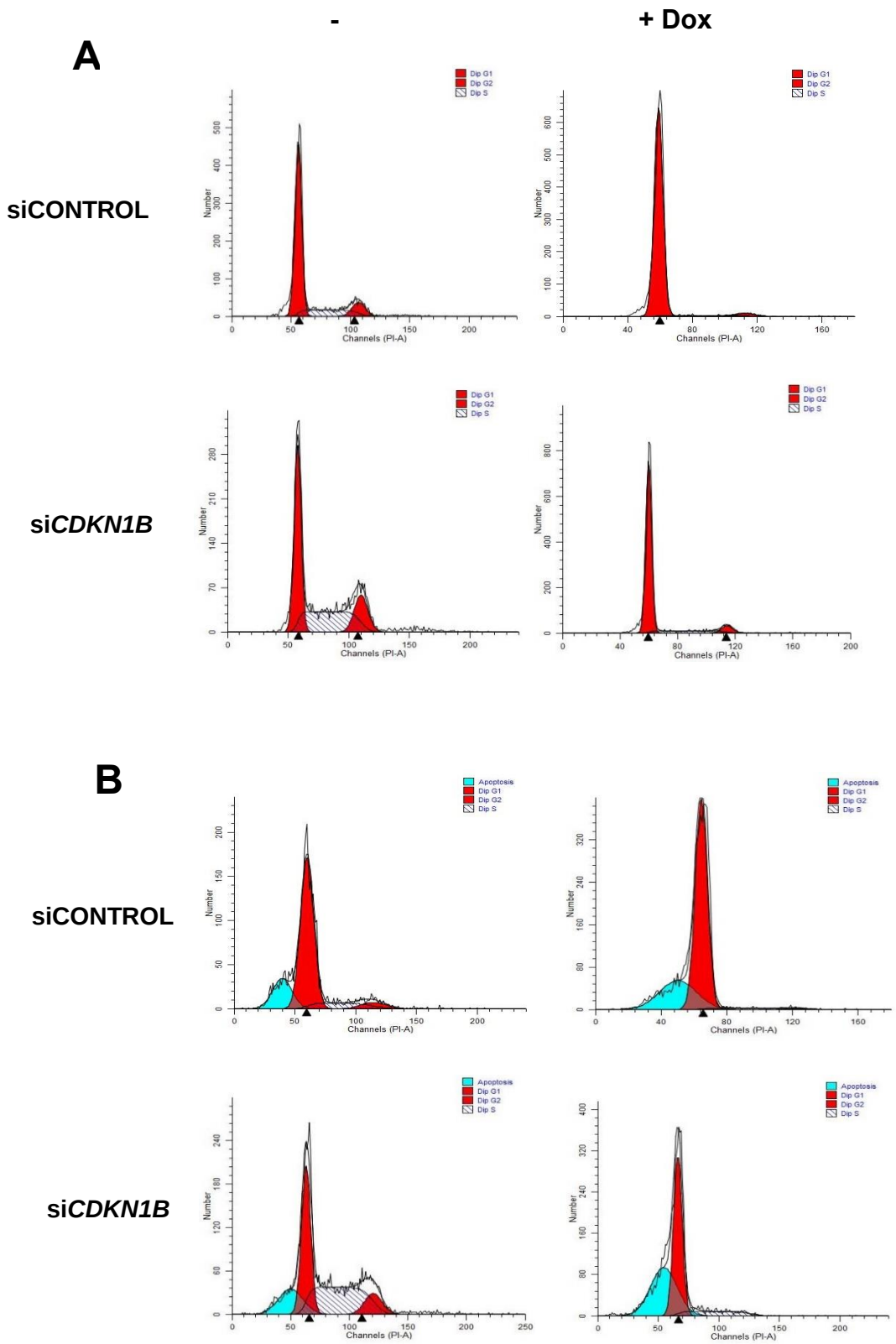


Figure 4.4: Loss of p27KIP1 expression abolishes the cell cycle arrest induced by ZEB1 overexpression.

A Analysis of cell proliferation after ZEB1 overexpression and p27Kip1 knockdown. Cells were transfected with either siCONTROL or si*CDKN1B* and left to grow overnight before seeding in Dox containing or deficient media. Cells were quantified at the indicated time points for both groups. Note that the kinetics of proliferation of ZEB1-expressing cells with depleted p27Kip1 (yellow line) is similar to that of control MCF7 cells (blue line). Data are presented as mean±SD. Each time point was quantified in triplicates and the growth curve was performed twice with similar results. **B** western blotting confirming p27Kip1 knockdown. Cells were transfected with either a control siRNA or si*CDKN1B* and cultured with or without Dox for 72h before lysing and analysis of the protein expression for the indicated genes. Tubulin was used as a loading control. This was repeated twice, and a representative blot is shown.

In addition, cell cycle distribution analysis of these cells showed a reduction of cells halted in G1 from 96% to ~83% in siCONTROL compared with si*CDKN1B* transfected cells (p-value <0.05). Similarly, an increase in S-phase from 2.6% to 12.7% was observed (p-value <0.05), but without a noticeable change in G2 (Figure 4.5).


-
+ Dox
A
siCONTROL
siCDKN1B
Dip G1
Dip G2
Dip S
Number
Channels (PI-A)
B
siCONTROL
siCDKN1B
Apoptosis
Dip G1
Dip G2
Dip S
Number
Channels (PI-A)

C

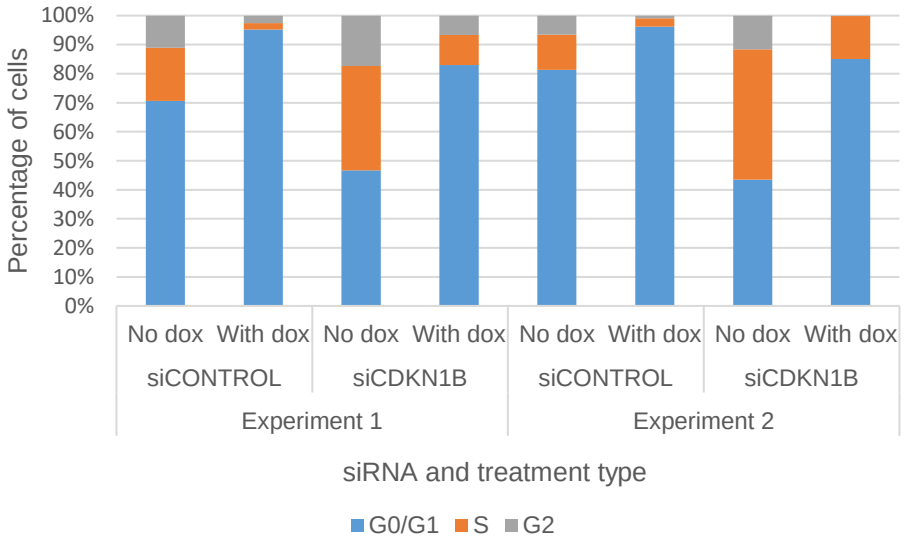


Figure 4.53: Loss of p27Kip1 expression alters cell cycle distribution of MCF7 cells overexpressing ZEB1 undergoing EMT.

Cells were transfected with either siCONTROL or si*CDKN1B* and cultured with/without 2µg/ml Dox before lysing for FACS analysis. A and B shows the shows the cell cycle distribution for two different experiments. C shows a graphical representation of the FACS analysis for the two experiments.

4.3.3 p27Kip1 is a stable protein in proliferating and resting cells

Two possible mechanisms could account for p27Kip1 upregulation, protein stability or mRNA level. p27Kip1 is degraded through the ubiquitin-proteasome pathway after its phosphorylation at Ser^{10} and export from the nucleus at the end of G1 phase permitting entry into S phase. The ubiquitination is induced by the E3 ligase RING finger and CHY zinc finger domain-containing protein 1, Pirh2 (Hattori et al., 2007). RNA sequencing data (collaboration with Dr Sayan, University of Southampton) has shown that overexpression of ZEB1 in MCF-7 cells down-regulated transcription of the *RCHY1* gene, a gene that encodes Pirh2 (approximately 4-fold, p-value <0.01 (0.0052), unpublished data). Accordingly, western blotting confirmed reduced expression of Pirh2 protein in ZEB1-expressing cells (Figure 4.6, A). So, we hypothesised that repression of Pirh2 led to ubiquitination and destabilisation of p27Kip1. To test this, cycloheximide assay was performed to assess the stability of the protein. I found that p27KIP1 was stable, even without ZEB1 overexpression (Figure 4.6, B). This suggests that p27Kip1 upregulation is not through its accelerated degradation.

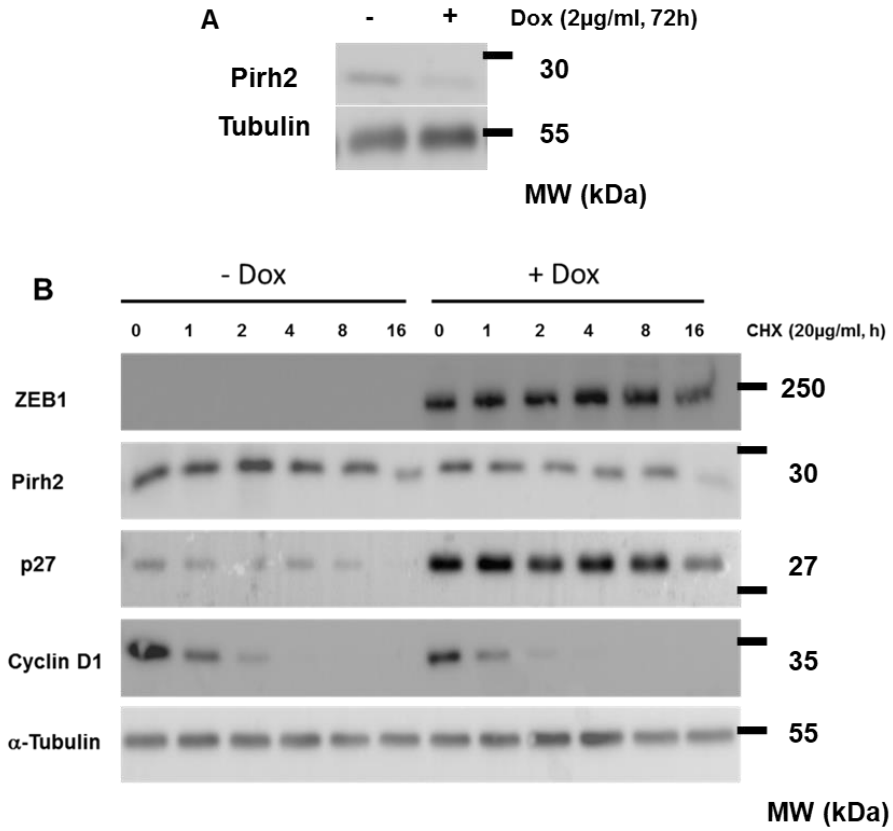


Figure 4.6: Pirh2 level is reduced upon ZEB1 overexpression and EMT induction (A), however, p27KIP1 is very stable (B).

A, Cells were cultured with or without DOX for 72 h before analysing for western blotting. B, Cells were cultured with or without DOX for 72 h before performing the cycloheximide assay and the cells were lysed for western blotting at the indicated time points. Cyclin D1 was used as a positive control for cycloheximide assay. Tubulin was used as a loading control. The experiment was repeated 3 times and a representative blot is shown.

4.3.4 p27Kip1 was upregulated at mRNA level.

qPCR was performed to study the effect of ZEB1 overexpression on the level of *CDKN1B* transcription in MCF7 cells. A significant increase of approximately 2.5-fold was observed in *CDKN1B* transcription after ZEB1 overexpression (i.e. EMT induction) (Figure 4.7).

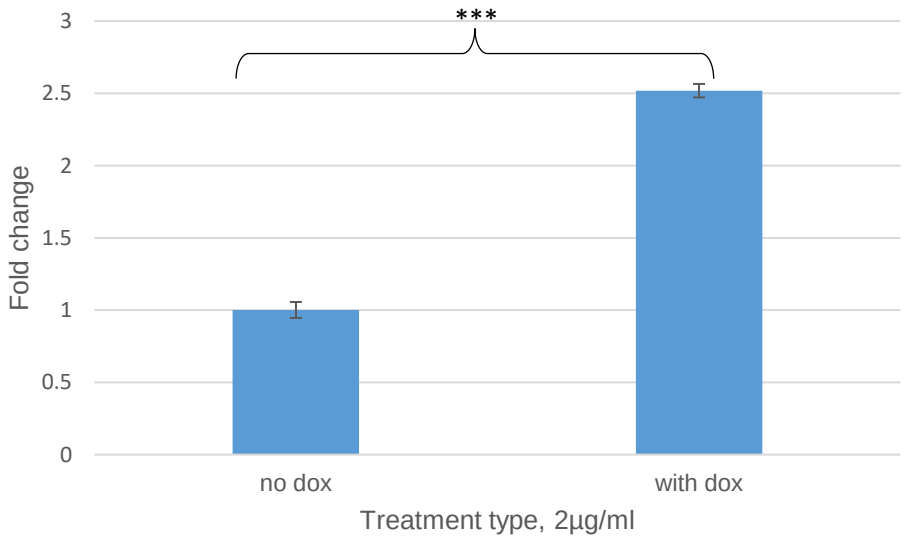


Figure 4.7: *CDKN1B* transcription level in MCF-7 ZEB1 (410) normally and following ZEB1 overexpression.

ZEB1 overexpression was found to increase the level of *CDKN1B* transcription. The 2-ΔΔCT method was utilised to analyse the qPCR data for *CDKN1B* level of expression. *** p-value <0.001. p-value = 0.0001. Data are plotted as mean±SD.

4.3.5 MicroRNA was not involved in differential expression of p27Kip1 protein in ZEB1-expressing and not expressing cells.

The mechanism by which ZEB1 overexpression upregulates p27Kip1 expression could be a result of differential expression of microRNAs (miRNAs) in ZEB1-expressing and non-expressing cells. The TargetScan database analysis

showed that the 3'UTR of *CDKN1B* is a target for several miRNAs including an evolutionary conserved binding site for miR200 family (Figure 4.8).

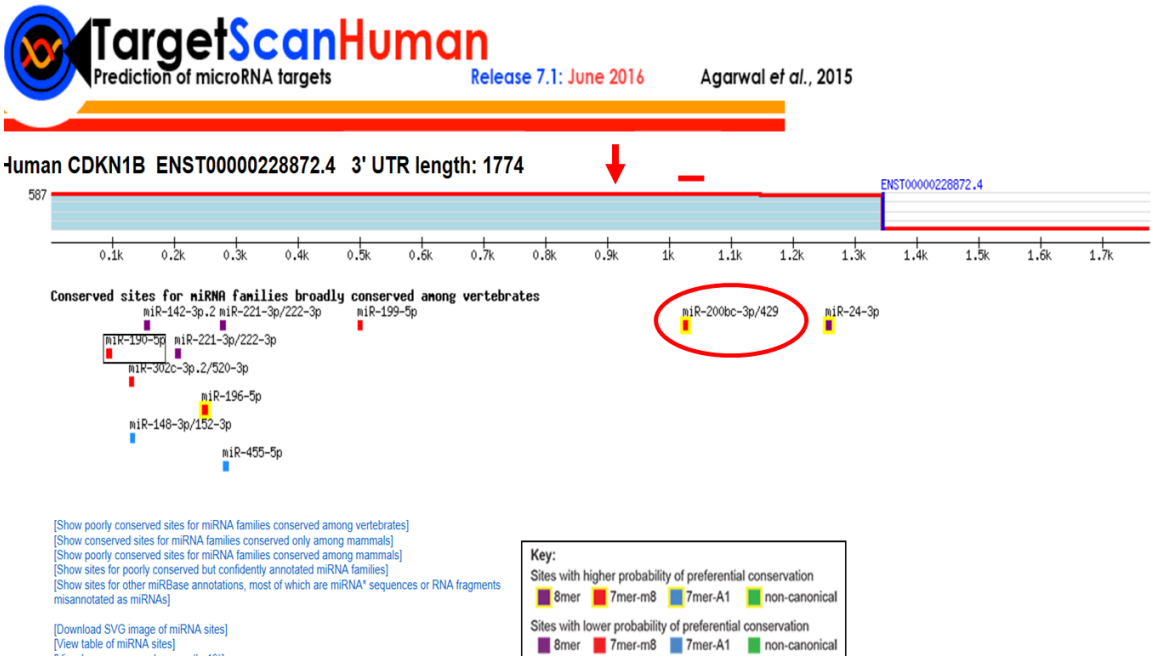


Figure 4.8: A map of the miRNAs that regulate the human *CDKN1B*.

miR200 family and its binding position (1022-1028) are highlighted with a red circle and dash. The red arrow highlights the restriction site for HindIII within the sequence, which creates the pMIR- CDKN1B-954.

miRNA sequencing data from our lab has shown that ZEB1 overexpression resulted in a statistically highly significant 8-fold reduction in the expression of miR-200c (Figure 4.9). To investigate implication of miR200 and the 3'UTR in *CDKN1B* regulation, two luciferase reporter plasmids were generated. In the first plasmid, pMIR- *CDKN1B*-954, only the proximal 954bp of *CDKN1B* 3'UTR, without the miR200c binding sequence, was inserted downstream of the

luciferase gene in pMIR-REPORT vector (appendix 4 & Figure 4.10). In the second plasmid, pMIR- *CDKN1B*-1400, the complete *CDKN1B* 3'UTR that contained the miR-200 binding site was cloned in the same plasmid (Figure 4.10).

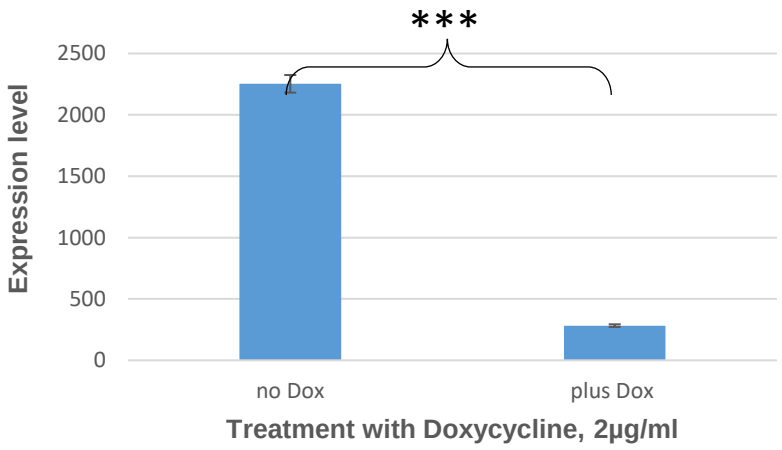


Figure 4.9: The reduction in the expression level of miR200c upon EMT induction by ZEB1 in MCF-7 cells.

RNA sequencing analysis was performed for cells cultured with/without 2µg/ml Dox for 72h. *** p-value <0.001. p-value = 0.0007.

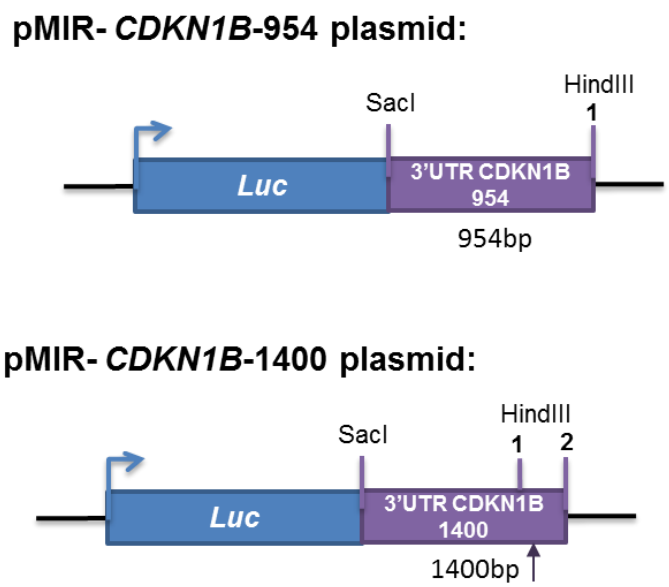


Figure 4.10: Schematic representation of the cloned plasmids for the 3'UTR of *CDKN1B*.

1 & 2 indicate the restriction sites for HindIII within the 3'UTR sequence. Arrow indicates the miR200 family binding position at 1022-1028.

4.3.5.1 Generation of pMIR- *CDKN1B*-954 and pMIR- *CDKN1B*-1400 plasmids

To generate the *CDKN1B* 3'UTR reporters, the 1400 bp insert was amplified using the OriGene pMirTarget-3'UTR -CDKN1B plasmid DNA as a template. The two inserts were then prepared with a complete (954 bp) and an incomplete (1400 bp) cleavage with HindIII followed by SacI cleavage. HindIII cuts around position 960 as well as at the end of the insert sequence producing three distinct bands, 1400bp, 954bp and 469bp (Figure 4.11).

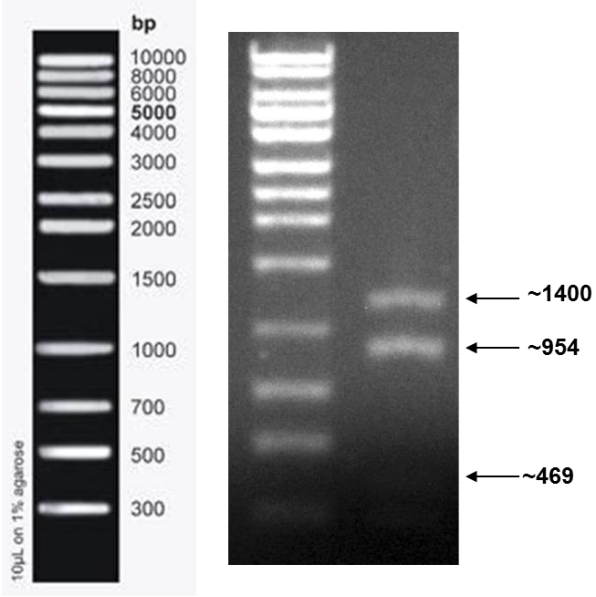


Figure 4.11: The 3 bands generated by SacI and HindIII restriction of the *CDKN1B* 3'UTR.

Lane 1 contains the marker, the HighRanger 1 kb DNA Ladder, and lane 2 contains the insert.

The 1400bp and the 954bp bands were excised from the agarose gel and purified prior to the cloning into the pMIR-REPORT vector between SacI and HindIII restriction sites. After transformation and final isolation, pMIR- *CDKN1B*-954 and pMIR- *CDKN1B*-1400 were restricted with SacI and HindIII to confirm the success of the cloning (Figure 4.12).

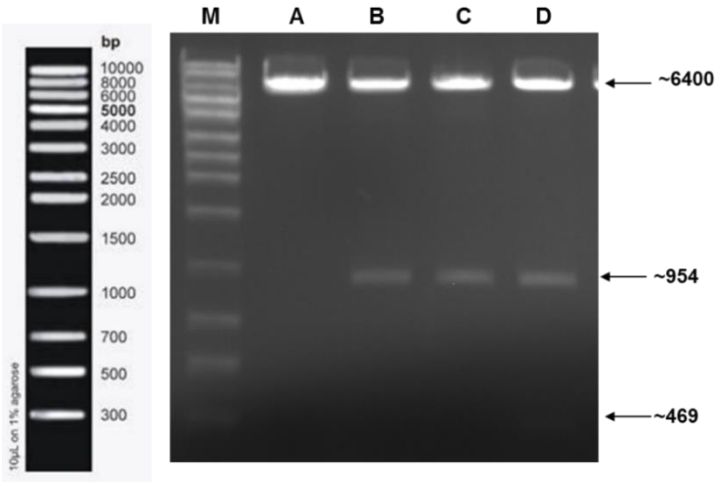


Figure 4.12: pMIR-*CDKN1B*-954 and pMIR- *CDKN1B*-1400 plasmids confirmation.

Samples of plasmid DNA were cleaved with SacI and HindII and analysed in the agarose gel electrophoresis. Lane A had the negative control, vector only, B and C show the pMIR-*CDKN1B*-954 where only the 954bp segment of the 3'UTR of *CDKN1B* is detectable as compared with the pMIR-*CDKN1B*-1400 plasmid that shows two bands (954bp and 469bp) of the complete *CDKN1B* 3'UTR (Lane D). M: the HighRanger 1 kb DNA Ladder.

Next, luciferase assays were performed to study the responsiveness of *CDKN1B* 3'-UTR to ZEB1 overexpression. If it was a target for miR200c, then the interaction between the 3'-UTR and the miRNA would result in the reduction of the luciferase activity and the detected signal. In addition, if miR200c was implicated, pMIR- *CDKN1B*-1400, but not pMIR-*CDKN1B*-954 reporter would show an increased luciferase activity in ZEB1-overexpressing cells, where miR200c was repressed. Transfection of the MCF-7 ZEB1 cells with the pMIR-*CDKN1B*-1400 plasmid was found to have a significantly reduced luciferase activity as compared to pMIR-*CDKN1B*-954 plasmid transfected cells, p-value <0.0001 (Figure 4.13). EMT induction was found to abolish the luciferase activity difference between the pMIR-*CDKN1B*-1400 and the pMIR-*CDKN1B*-954 plasmid transfected cells, p-value >0.05.

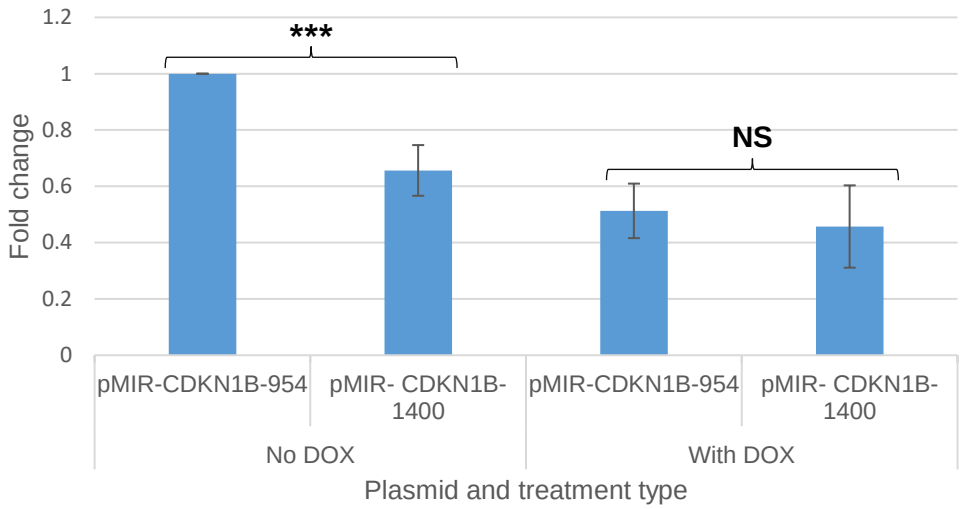


Figure 4.13: The activity of the pMIR-*CDKN1B*-954 and pMIR-*CDKN1B*-1400 luciferase reporters in MCF-7/ZEB1 cells before and after ZEB1 overexpression (i.e. EMT induction).

Cells were transfected with either pMIR-*CDKN1B*-954 or pMIR-*CDKN1B*-1400 along with a β-galactosidase reporter plasmid and cultured with or without Dox for 72h. The pMIR-*CDKN1B*-954 contains the 3'UTR of *CDKN1B* lacking the miR200c binding site and is used as a control, while the pMIR-*CDKN1B*-1400 contains the full 3'UTR of *CDKN1B* and is used to determine whether *CDKN1B* is a target for miR200c. β-galactosidase assay was used for luciferase activity normalisation. ******* p-value <0.001. **NS** non-significant. Data are plotted as mean±SD. N=3.

Next, a luciferase reporter was generated to investigate an effect of ZEB1 on *CDKN1B* promoter activity. In this reporter plasmid, the *CDKN1B* promoter was inserted upstream of the luciferase gene in the pGL3 basic plasmid (Figure 4.14).

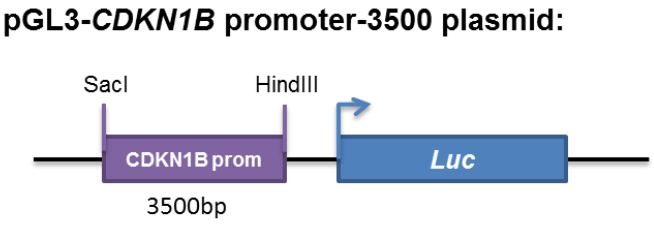


Figure 4.14: Schematic representation of the reporter containing the 3.5 kb *CDKN1B* promoter sequence.

4.3.5.2 Generation of pGL3-*CDKN1B* promoter-3500

Two steps amplification was performed to amplify the whole 3500bp sequence located upstream of *CDKN1B* gene using genomic DNA as a template. The two overlapping DNA fragments were amplified and mixed then used as a template to amplify the whole 3500bp sequence with the primers containing SacI and HindIII restriction sites used for cloning in pGL3 reporter vector (see materials and methods for details). To confirm the success of the cloning recombinant DNA samples were cleaved with SacI and HindIII and analysed in agarose gel electrophoresis (Figure 4.15).

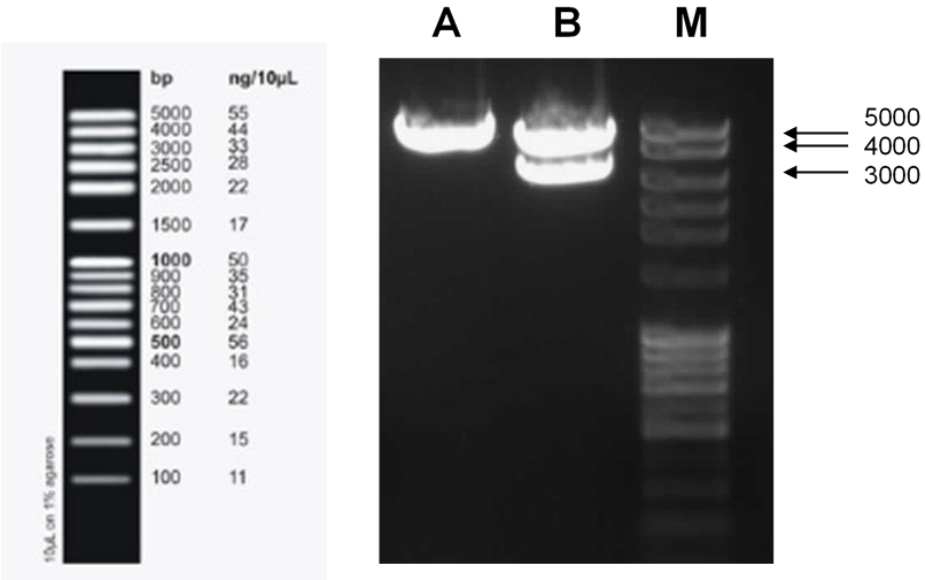


Figure 4.15: pGL3-*CDKN1B* promoter-3500 plasmid confirmation.
Samples of plasmid DNA were cleaved with SacI and HindIII and analysed in 1% agarose gel electrophoresis. Lane A: negative control (pGL3-basic). Lane B: pGL3-*CDKN1B* promoter-3500 plasmid). M: the FullRanger 100bp DNA Ladder.

Transient transfection of MCF7/ZEB1 cells with this reporter construct has demonstrated that ZEB1 overexpression significantly activated *CDKN1B* promoter, around 2-fold increase. Activity of the SV40 promoter reporter used as a negative control was not affected (Figure 4.16). This suggests that the upregulation of p27Kip1 expression induced by ZEB1 occurs at the level of transcription initiation.

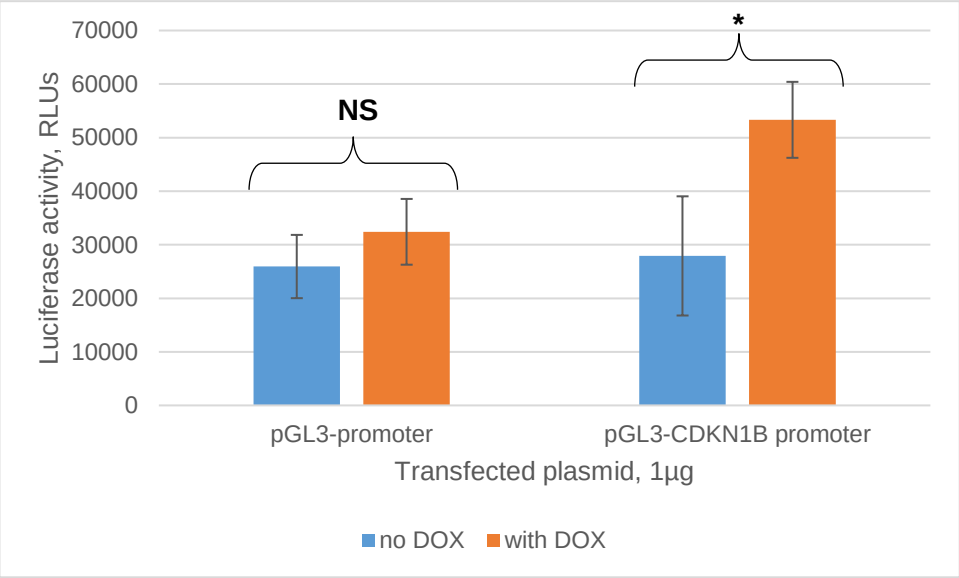


Figure 4.16: The activity of luciferase reporters pGL3-promoter and pGL3-*CDKN1B* promoter-3500 in MCF-7 ZEB1 following EMT induction by ZEB1.

Cells were transfected with either pGL3-promoter or pGL3-*CDKN1B* promoter-3500 along with a β-galactosidase reporter plasmid and cultured for 72h in the presence or absence of 2µg/ml DOX. This was to determine the effect of ZEB proteins overexpression on the activity of the *CDKN1B* promoter. pGL3-promoter contains SV40 promoter and is used as a negative control. β-galactosidase assay was used for luciferase activity normalisation. **NS** non-significant. ***** p-value <0.05 (0.029). Data are plotted as mean±SD. N=3.

4.3.6 Loss of p27KIP1 expression results in proliferating ZEB1-overexpressing cells deficient in LIG1

After observing that p27Kip1 rescues ZEB1-induced cell cycle arrest, I examined how that would affect Rb, cyclin D1 and LIG1 expression. I found that ZEB1-mediated reduction in cyclin D1 and LIG1 expression and Rb hypophosphorylation were not lost in p27Kip1-depleted cells (Figure 4.17). This suggests that loss of p27Kip1 produces proliferating ZEB1 ovrexpressing cells that are deficient in LIG1.

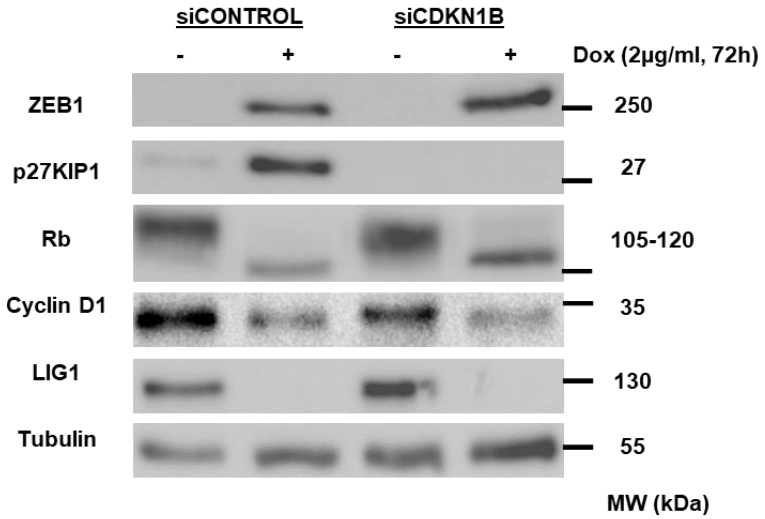


Figure 4.17: p27KIP1 Knockdown does not reverse the reduction in cyclin D1 and LIG1 expression or Rb hypophosphorylation induced by ZEB1-overexpression in MCF7 cells.

Cells were transfected with either siCONTROL or siCDKN1B and cultured in media with/without Dox for 72h before analysing with western blotting. Upper band of Rb= hyperphosphorylated, Lower band in Rb staining= hypophosphorylated. Tubulin was used as a loading control. The experiment was repeated 3 times and a representative blot is shown.

4.4 Discussion

This chapter aimed to identify whether ZEB1 overexpression (i.e. EMT induction) in MCF-7 cells would result in cell cycle arrest as previously shown for other cell models of EMT. If it led to cell cycle arrest, the aim was then to identify the mechanism causing this arrest. Here we show that EMT induction in MCF-7 cells by ZEB1 also resulted in cell cycle arrest in G1 phase (Figure 4.1). As demonstrated using RNAi, this arrest was mediated by the upregulation of the CDK-inhibitor p27Kip1 (Figure 4.2). Promoter reporter assays demonstrated that ZEB1 overexpression enhanced *CDKN1B* promoter activity. It remains unclear whether this effect is direct or not. Employing chromatin immunoprecipitation method in further experiments could clarify this issue. Cells in the EMT state after ZEB1 overexpression with depleted p27Kip1 re-entered cell cycle, but displayed typically mesenchymal morphology (Figures 4.3 and 4.4).

So, our finding here suggests that p27Kip1 has a tumour suppressive and protective role during EMT in MCF-7 cells. The prime function of p27Kip1 is in the inhibition of cyclin D-cdk4 and cyclin E-cdk2 complexes and hence prevention of G1-S cell cycle transition. *CDKN1B* mutations in human cancers has rarely been documented, however, whenever there was a loss of p27Kip1 function, that was associated with poor prognosis and high histopathological grades of tumours (Chu et al., 2008). Reduced expression of p27Kip1 was often associated with the increased expression of its degrading ubiquitin ligases such as Skp2 leading to high proliferative potential (Chen et al., 2014; Lv et al., 2017). In addition, a recent study has suggested that *CDKN1B* was one of the most mutated genes in luminal A breast cancer. Most of these mutations affected the C-terminal region of the protein, which harbours the protein stability moiety (Belletti and Baldassarre, 2012; Stephens et al., 2012). A Germ-line mutation of *Cdkn1b*, a homozygous frameshift, has been shown to result in a strong reduction in p27Kip1 levels in MENX-affected rats. Additionally, a patient with a MEN1 negative multiple endocrine neoplasia-like syndrome (MENX) was found to harbour a nonsense germ-line mutation of *CDKN1B* that led to pituitary and parathyroid tumours. This suggests that multiple endocrine tumours could be predisposed by germ-line mutations in *CDKN1B* (Pellegata et al. 2006).

p27Kip1 plays a critical in role maintaining quiescence and terminal differentiation of adult cells. The deletion of p27Kip1 in postnatal mice was shown to be sufficient for the organ of corti and pars intermedia proliferation induction. Additionally, the expression of p27Kip1 in animals that have pituitary tumours was able to cease tumour growth, which suggest the ability of p27Kip1 to induce growth inhibition even at adult age (Oesterle et al., 2011).

Accumulating evidence suggests that p27Kip1 has a functional duality in tumours. While p27Kip1 original cell cycle dependent role has been implicated with tumour suppressive mode of action, recent studies has suggested cell cycle independent roles that are oncogenic in nature (Sicinski et al., 2007). For instance, p27Kip1 mislocalisation to the cytoplasm has been suggested to activate cellular motility and invasion contributing to tumour progression (Besson et al., 2004a). In addition, knock-in mouse with 4 amino acid substitution in p27Kip1 ($p27^{Ck-}$), which impedes its interaction with CDK/cyclins, led to multiple hyperplastic lesions and tumours in many organs including lung and retina. This was due to the amplification of stem/progenitor cells. $p27^{Ck-}$ has also been shown to be mislocalised to the cytoplasm suggesting that cytoplasmic p27Kip1 acts as an oncoprotein (Besson et al., 2007).

We have also investigated the mechanism by which ZEB1 expression upregulates p27Kip1 expression. This regulation could either be at the protein level, such as degradation inhibition, or at RNA level, such as miRNA inhibition or direct promoter interaction. RNA sequencing data from our lab has shown that ZEB1 expression in MCF-7 cells resulted in a statistically significant 4-fold decrease in the transcription of *RCHY1*, which encodes for Pirh2, an E3 ligase implicated in p27Kip1 ubiquitination (p-value <0.01 (p-value< 0.01, 0.0052), unpublished data). Although Pirh2 protein was found to be reduced upon ZEB1 overexpression in MCF-7/ZEB1 cells (Figure 4.6, A), cycloheximide assay has confirmed the stability of the protein even without ZEB1 overexpression (Figure 4.6, B).

After that, we investigated a potential effect at the RNA level. qPCR demonstrated a statistically significant increase of approximately 2.5-fold in the level of *CDKN1B* transcription after ZEB1 overexpression in MCF-7/ZEB1 cells

(Figure 4.7). So, initially, we examined the potential interference with miRNA induced inhibition. TargetScan database identified the 3'UTR of *CDKN1B* as a target for various miRNAs including miR200c (Figure 4.8), which was found to be reduced upon by ZEB1 (Figure 4.9). Transfection of the MCF-7/ZEB1 cells with the pMIR-*CDKN1B*-1400 plasmid, which harbours the miR200c binding site, was found to have a significantly reduced luciferase activity as compared to pMIR-*CDKN1B*-954 plasmid, which lacks the miR200c binding site, transfected cells, p-value <0.0001 (Figure 4.13). This implicates *CDKN1B* as a miR200c target. However, ZEB1 overexpression (i.e. miR200c reduction) was merely found to abolish the luciferase activity difference between the pMIR-*CDKN1B*-1400 and the pMIR-*CDKN1B*-954 plasmid transfected cells, p-value >0.05. This did not result in an increase luciferase activity of the pMIR-*CDKN1B*-1400 transfected cells as expected. This could implicate *CDKN1B* 3'UTR in this system as a target for another miRNA that is not affected by ZEB1. Another example of miRNAs that regulate *CDKN1B* 3'UTR that was identified by TargetScan is miR200b, which was not found to be downregulated by ZEB1 overexpression in the miRNA sequencing data (unpublished data). In fact, p27Kip1 was shown to be a target for miR200b in human Tenon's capsule fibriblasts (Tong et al., 2014). More investigations should be performed to confirm whether p27Kip1 is a target of miR200b.

In addition, luciferase assays demonstrated a statistically significant increase of ~2-fold in the activity of *CDKN1B* promoter as compared to that of the SV40 promoter as a result of ZEB1 expression (Figure 4.16). This proposes a direct interaction between ZEB1 and the *CDKN1B* promoter. More studies are required to confirm this link.

Here I found that p27Kip1 knockdown uncouples EMT induction from other effects of ZEB1 overexpression (i.e. quiescence), but produced no effect on Rb phosphorylation, cyclin D1 and LIG1 expression levels (Figure 4.17). So, loss of p27Kip would generate cells that are undergoing EMT whilst LIG1 deficient. This could predispose those cells to genomic instability due to reduced DNA repair efficiency.

p27Kip1 deficiency in luminal breast cancer resulted in genomic instability as well as radioresistance . p27Kip1-deficient cells accumulated residual DNA damage, mitotic defects and genetic instability, but this was not linked with enhanced cytotoxicity, suggesting that these genetically unstable cells have developed some form of survival mechanisms (Berton et al., 2017).

Chapter 5 : Transient ZEB1 overexpression and LIG1 reduction results in an increase in DNA damage and chromosomal aberrations

5.1 Introduction

LIG1 is involved in BER, NER, HR as well as in the ligation of Okazaki fragments of the lagging strand during DNA replication. Hence, loss of LIG1 in proliferating cells would be expected to result in the accumulation of DNA damage. In fact, the deficiency of LIG1 in 46BR1G1 cells has been shown to result in replicative defects characterised by the accumulation of SSBs as well as DSBs, which may potentially lead to genomic instability. (Soza et al., 2009).

G1 phase cell cycle arrest implied by EMT-TFs could serve as a protective mechanism preventing DNA damage in S phase and chromosomal abnormalities in mitosis. We have shown that this arrest was due to an upregulation of p27Kip1 tumour suppressor, as its knockdown was enough to revoke the quiescence. However, loss of p27Kip1 was not sufficient to restore LIG1 expression, and p27Kip1 KD cells entered cell cycle being LIG1-deficient. As EMT is a reversible process in many cancer types, we proposed that EMT-TFs-mediated cell cycle exits and re-entries would impact on genome stability due to the reduced LIG1 levels.

In this chapter, we modelled stable and transient ZEB1 overexpression (i.e. EMT) by p27Kip1 KD and by sequentially incubating MCF-7/ZEB1 cells in Dox-containing and Dox-free media. In transient ZEB1 overexpression model, we have demonstrated that the cells start accumulating DNA damage, as they enter the cell cycle following Dox withdrawal. Karyotype analysis has demonstrated that stable ZEB1 overexpression in proliferating cells, and transient ZEB1 overexpression affect chromosomal integrity and produce aberrations. Moreover, *LIG1* knockdown in proliferating MCF-7 cells was sufficient to significantly increase the levels of chromatid breaks and fusions. Those results confirm that both stable and transient ZEB1 overexpression can result in genomic instability, and reduced levels of LIG1 is a contributing factor. As neither LIG3 or LIG4 were not affected by ZEB1 overexpression, we conclude that LIG1 is not redundant in this model.

5.2 Aims

- Identify the effect of EMT induction and LIG1 loss on DNA damage in MCF-7 ZEB1 cells

- Perform karyotyping of metaphase spreads to examine the effect of EMT and LIG1 loss on chromosomal integrity

5.3 Results

5.3.1 LIG1 knockdown in MCF-7 cells results in an increase in chromatid breaks and chromosomal/chromatid fusions

We found that overexpression of ZEB1 in conjunction with p27Kip1 loss led to proliferating cells with mesenchymal morphology and reduced levels of LIG1 (Figures 4.3, 4.4 and 4.17). As LIG1 is essential for DNA repair and DNA replication, we proposed that its loss is a factor impacting on genome integrity. To test this, I employed a method of chromosome karyotyping using metaphase chromosome spreads. To validate the methodology, I made use of phleomycin, a genotoxic agent whose toxicity is achieved by the distortion of the DNA double helix integrity through intercalating the DNA. Phleomycin induces predominantly chromatid breaks, and some other aberrations, albeit with much lower frequencies. MCF-7 were treated with 25µg/ml of phleomycin, or left untreated, then, the cells were exposed to colcemid to enrich cell populations for metaphase cells and analysed for chromosomal aberrations (Figure 5.1, A & B). Expectedly, I found a 25-100-fold increase in the incidences of chromatid breaks in different experiments, but other phleomycin-induced aberrations were less frequent (Figure 5.2). Indeed, I found 1.6- and 6.3-fold phleomycin-induced increase in chromosomal breaks and fusions, respectively.

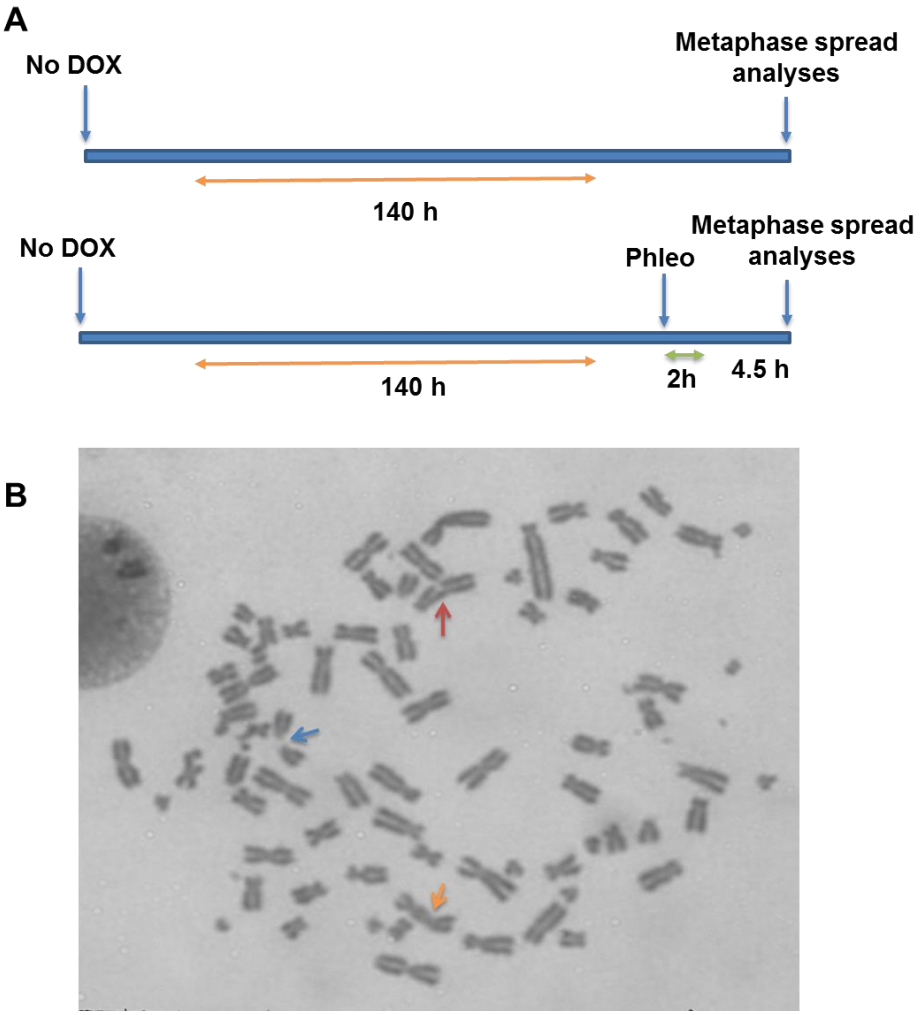


Figure 5.1: A scheme of phleomycin karyotyping experiment

showing the duration of the experiment and at which stage the metaphase spread analyses were performed, **A**. **B** an example of a metaphase spread showing the types of chromosomal aberrations scored, Red arrow, chromatid break, Blue Arrow, chromosome break, Orange arrow, fusions. More examples of fusions are shown in Appendix 8.

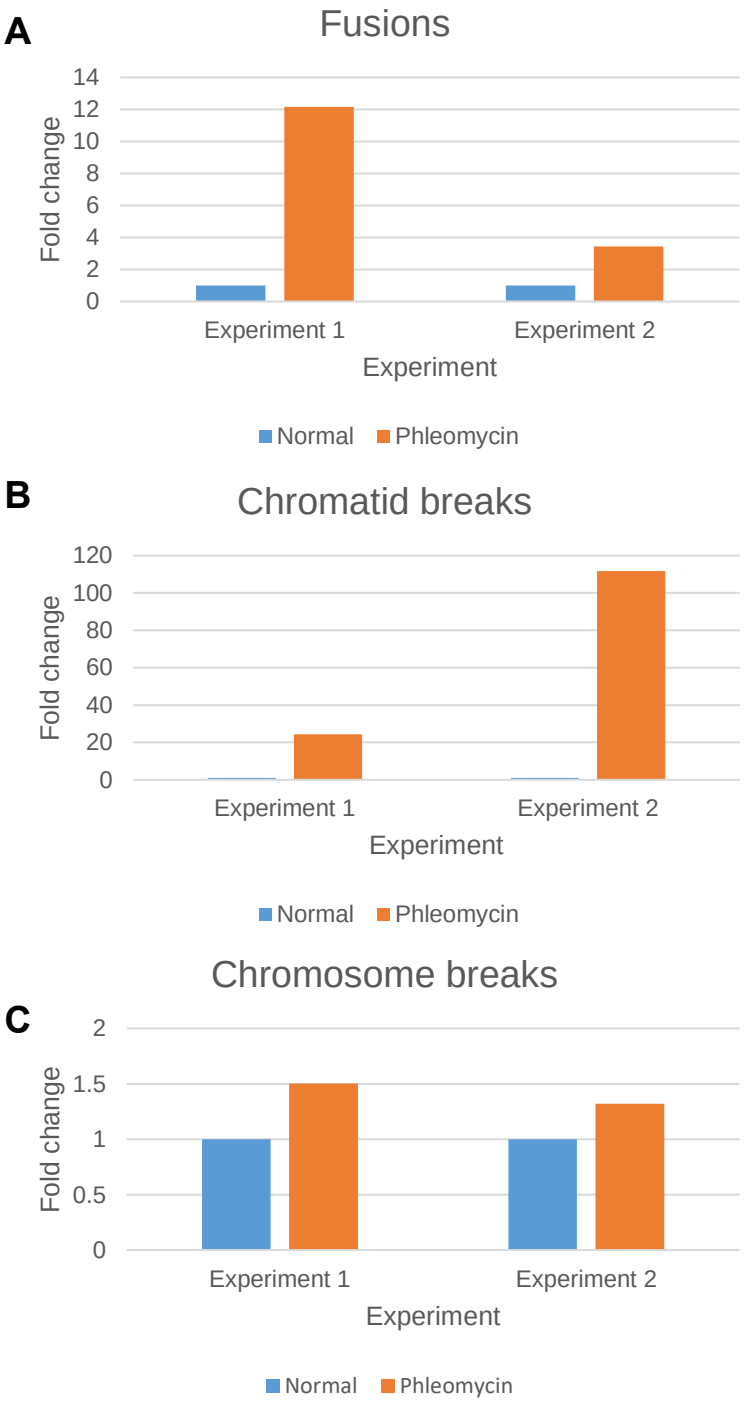


Figure 5.2: Phleomycin treatment of MCF-7 cells induces chromosomal aberrations, especially chromatid breaks.

The figure shows an increase in chromatid breaks and fusions in MCF-7 ZEB1 cells as the result of phleomycin treatment as compared to normal cells with a slight increase in chromosomal breaks. n=2. Number of spreads=35.

MCF-7/ZEB1 cells were transfected with *LIG1*-targeting or control siRNAs; cells were maintained in Dox-free media, arrested in metaphase phase by exposure to colcemid for 1 hours, and spreads were prepared and analysed microscopically as described in section 2.2.11 (Figure 5.3). Transfecting cells with si*LIG1* was shown to increase the level of chromatid breaks and fusions by 2.3 and 4.8-folds respectively as compared to the cells transfected with control siRNA (Figure 5.4, A). Increases in the level of chromatid breaks and chromosomal/chromatid fusions were statistically significant (p-value <0.05) and very statistically significant (p-value <0.01), respectively. The incidence of chromosomal breaks was the same in control and *LIG1* knockdown cells, as the control MCF-7 cells displayed a relatively high level of chromosomal breaks.

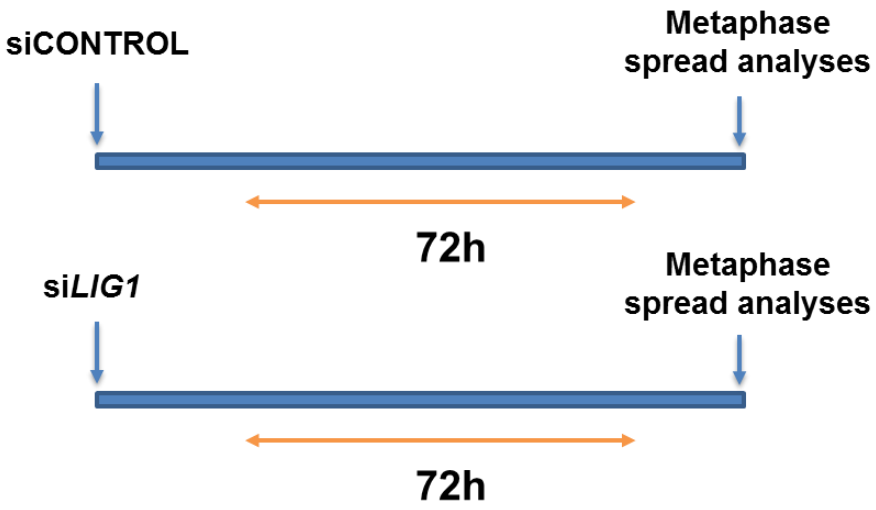


Figure 5.3: Metaphase spread analyses to investigate the effect of *LIG1* loss on chromosomal integrity.

a scheme of the LIG1 knockdown karyotyping experiment showing the duration of the experiment and at which stage the metaphase spread analyses were performed.

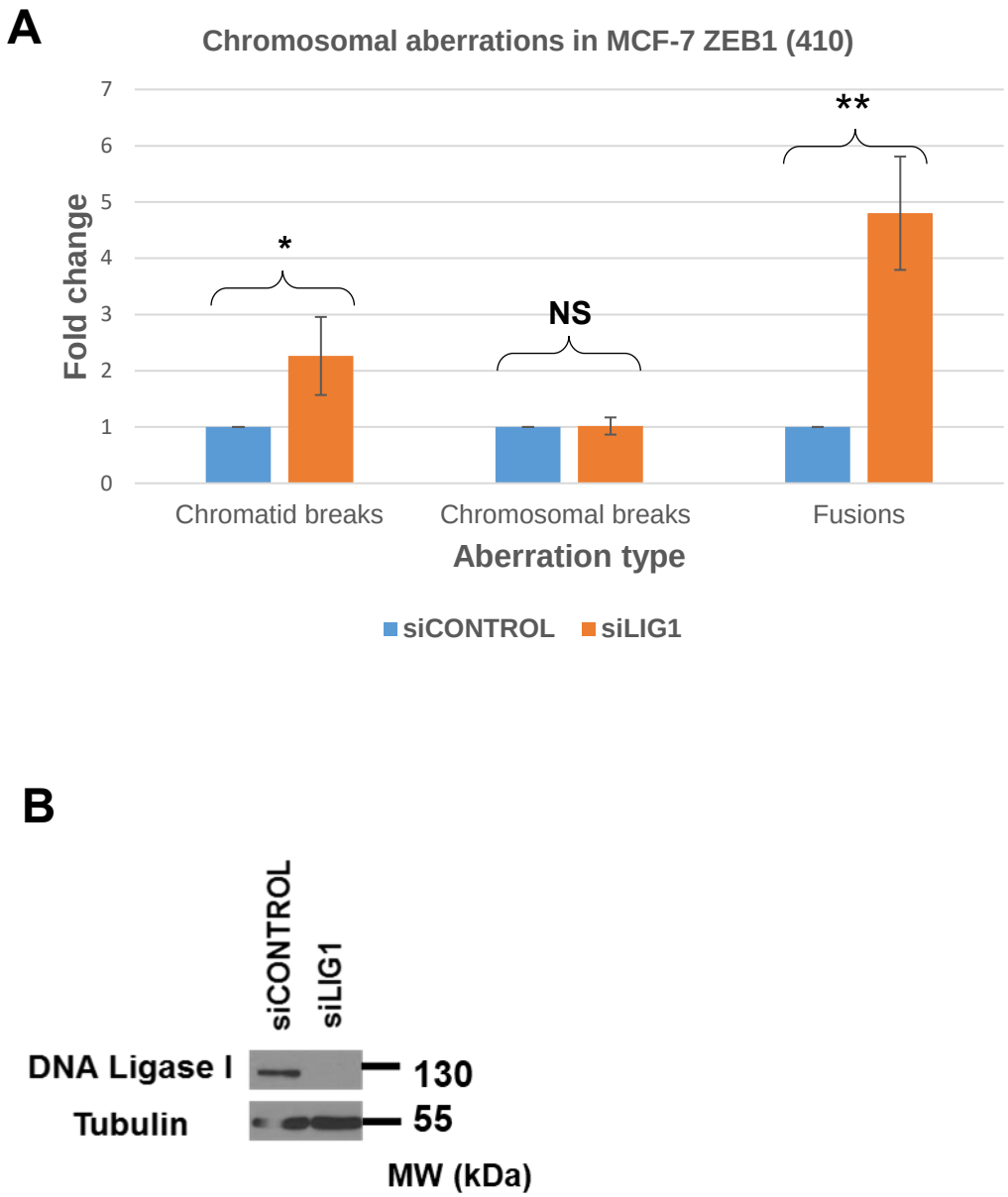


Figure 5.4: LIG1 knockdown induces chromosomal aberrations.

The figure shows an increase in chromatid breaks and fusions in MCF-7 ZEB1 cells after LIG1 knockdown (siLIG1) as compared to control (siCONTROL) **(A)** western blotting confirming DNA Ligase I knockdown **(B).** * p-value <0.05 (0.0324). ** p-value <0.01 (0.0026). **NS** non-significant, p-value >0.05 (0.75). n=3. Number of spreads=35.

5.3.2 Stable ZEB1 overexpression coupled with p27Kip1 knockdown increases chromosomal/chromatid fusions

LIG1 depletion significantly increased the incidence of chromatid breaks and fusions (Figure 5.4), and depletion of p27Kip1 results in ZEB1-expressing LIG1-deficient proliferating cells (Figure 4.17). Therefore, we reasoned that ZEB1-overexpression concomitant with p27Kip1 loss may lead to chromosomal instability. To test this hypothesis, we performed karyotype analysis of MCF-7/ZEB1 cells with depleted p27Kip1 representing a model of a stable ZEB1 overexpression (i.e. EMT). Cells were transfected with either siCONTROL or si*CDKN1B* and maintained in the presence of absence Dox for 72h, then metaphase spreads were prepared and analysed as described in Materials and Methods section (Figure 5.5).

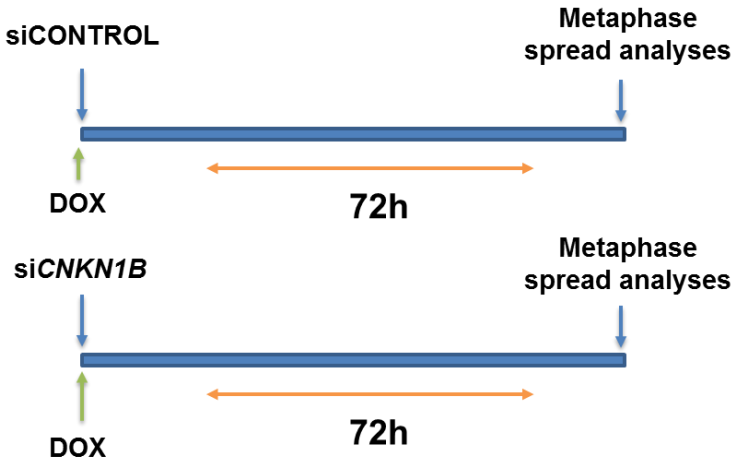


Figure 5.5: A scheme of the *CDKN1B* knockdown and ZEB1 overexpression karyotyping experiment

showing the duration of the experiment and at which stage the metaphase spread analyses were performed.

ZEB1 overexpression concomitant with p27Kip1 loss caused statistically significant increase in chromosomal/chromatid fusions to approximately 6-fold as compared to the MCF-7 cells with depleted p27Kip1 but grown in the Dox-free media (Figure 5.6, A). The amount of chromatid breaks was approximately 2.4-fold higher in ZEB1-overexpressing cells than in epithelial control (Figure 5.6, A). However, this difference did not reach statistical significance (p-value >0.05). Thus, the effect of ZEB1 overexpression on chromosomal aberrations in this model was very similar to that observed in MCF-7 cells with RNAi-mediated depletion of LIG1.

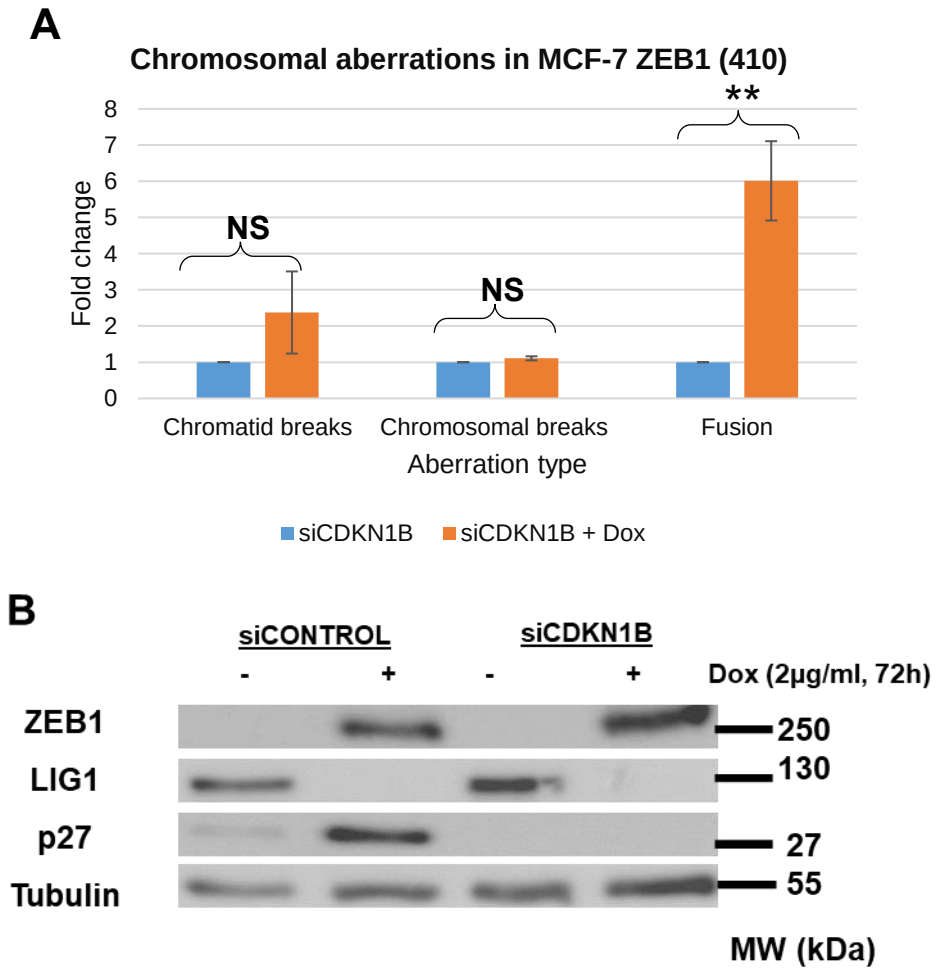


Figure 5.6: *CDKN1B* knockdown concomitant with ZEB1 overexpression in MCF-7 cells induces chromosomal aberrations, especially fusions.

The figure shows an increase in chromatid breaks and fusions in MCF-7 ZEB1 cells as the result of *CDKN1B* knockdown coupled to EMT induction as compared to control (si*CDKN1B* but no dox) (**A**) western blotting confirming p27Kip1 knockdown, ZEB1 expression and LIG1 deficiency. (**B**). ****** p-value <0.01 (0.0015). **NS** non- significant, p-value >0.05 (chromatid breaks: 0.10, chromosomal breaks: 0.13). n=3. Number of spreads=35.

5.3.3 Transient ZEB1 overexpression results in DNA damage

EMT is a reversible process in embryogenesis, and it has been proposed that reversible EMT programs prevail in most epithelial cancers (Nieto et al., 2016). We hypothesised that transient EMT may impact on genome integrity in cancer cells. To this end, we performed experiments aiming: i) to establish whether manipulations with Dox in the media could be used to model transient EMT in vitro, and ii) to analyse how transient EMT would affect endogenous DNA damage, and the expression levels of cell cycle markers and DNA ligases. So, EMT was induced by ZEB1 overexpression in MCF-7 cells for 72h and then Dox was removed, and cells were collected at different time points. Dox withdrawal resulted in gradual decrease in the expression levels of ZEB1, which became undetectable in cells grown in Dox-free medium for 48 hours. Concomitant with loss of ZEB1, expression level of p27Kip1 also decreased to the level observed in Dox-untreated cells. At time point 72 hours, the expression level of LIG1 was restored, but ZEB1 overexpression (i.e. EMT) seems to have a minimal if any effect on LIGIII and no effect on LIGIV. Immunoblotting for phosphorylated histone H2A.X (ɣ-H2A.X) was performed to detect DNA damage. Upon the exposure of cells to endogenous or exogenous DNA-damaging agents, they accumulate DSBs causing phosphorylation of the histone variant H2A.X at Ser139 in its carboxyl tail domain. The level of histone phosphorylation is representative of the level of the DSBs, and for this reason, ɣ-H2A.X is considered as a marker of DNA damage (KUO and YANG, 2008). A gradual increase in ɣ-H2A.X levels was observed following Dox withdrawal with a peak after around 24h (Figure 5.7). Interestingly, although both other two ligases are present in the cells, they do not seem to compensate for LIG1 loss to prevent endogenous DNA damage.

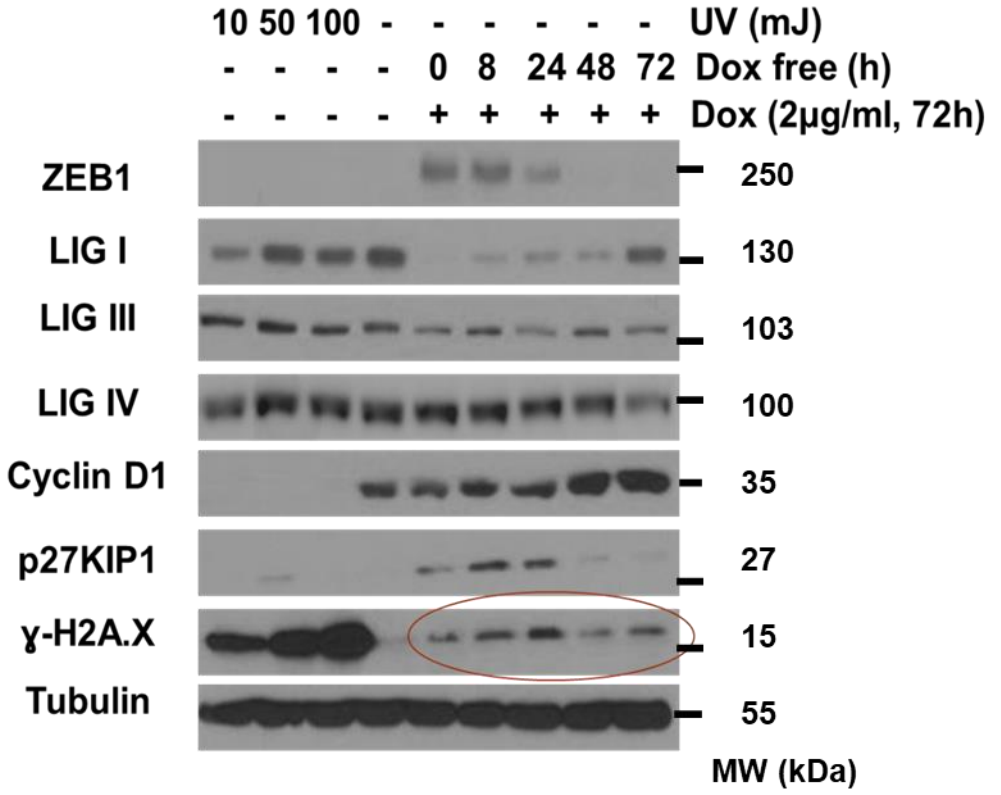


Figure 5.7: Transient ZEB1 overexpression resulted in an increase in the level of DNA damage.

ZEB1 and p27Kip1 expression level gradually decreases after Dox withdrawal concomitant with an increase in LIG1. LIGIII and LIGIV were either minimally affected or not affected, respectively. γ-H2A.X levels gradually increase following Dox treatment (i.e. EMT induction) as well as withdrawal with a peak at 24h. UV treatment is used as a positive control for γ-H2A.X. Tubulin is used as a loading control. The experiment was repeated 3 times and a representative blot is shown.

5.3.4 Transient ZEB1 overexpression in MCF-7 cells also results in an increase chromosomal aberration

As transient ZEB1 overexpression is accompanied by DNA damage and transient loss of LIG1, I performed karyotype analysis to investigate the effect of transient ZEB1 overexpression on chromosomal integrity in MCF-7 cells. Here, cells were treated with Dox for 72h and then maintained in the absence or presence of Dox for additional 72 hours (Figure 5.8). Interestingly, as in the experiments with *LIG1* KD (subchapter 5.3.1) or in a cellular model of stable ZEB1 overexpression (subchapter 5.3.2), the most significant difference was observed in chromosomal/chromatid fusions (5.6-fold increase, p-value <0.05, 0.049) (Figure 5.9). Chromatid and chromosomal breaks displayed 2.5- and 1.3-fold difference, respectively, compared to uninduced cells (Figure 5.9). However, these results were statistically insignificant (p-value >0.05, p-values were 0.1084 and 0.105, respectively).

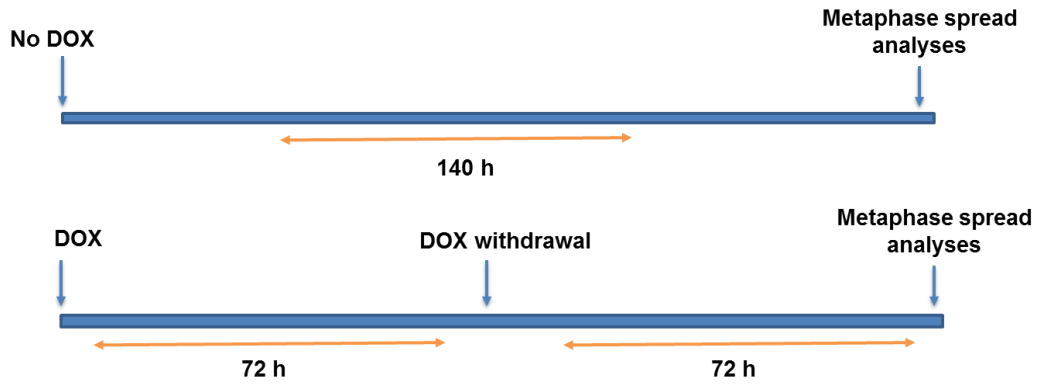


Figure 5.8: A scheme of the transient ZEB1 expression karyotyping experiments

showing the duration and the stages of the experiments and at which stage the metaphase spread analyses were performed. Dox treatment was 2µg/ml.

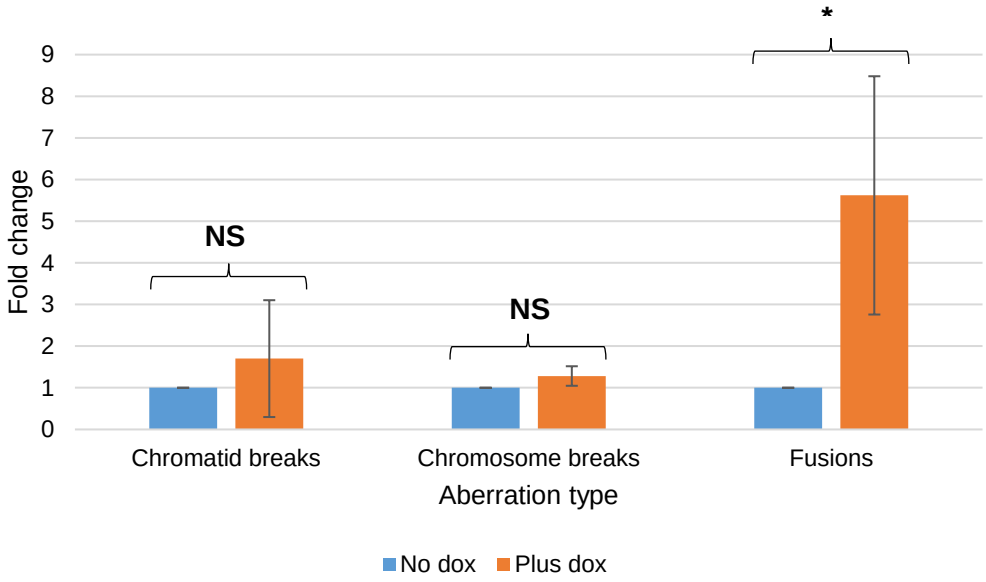


Figure 5.9: Transient ZEB1 overexpression in MCF-7 ZEB1 (410) cells results in an increase in the levels chromosomal aberrations.

An increase was observed in fusions and chromatid breaks but only a minimal increase in chromosomal breaks as compared to uninduced cells. Dox treatment was 2µg/ml. ***** p-value <0.05. **NS** non-significant. n=3. Number of spreads=35.

5.3.5 EMT induction in a stable A431-ZEB2-cyclin D1 cell model also results in an increase in fusions

All the experiments presented in this chapter so far were carried out in MCF-7 cell line. To generalise our findings and to test whether our observations are relevant to different cell lines, we employed another cellular model of stable EMT. Concomitant Dox-regulated expression of ZEB2 and cyclin D1 in A431 squamous carcinoma cells results in an EMT, but it does not affect cell cycle progression (Mejlvang et al., 2007). Thus, A431/ZEB2/Cyclin D1 cells were treated with Dox for 72h or left untreated prior to performing the metaphase spreads analyses (Figure 5.10). It was observed that ZEB2 overexpression in these cells also resulted in a statistically significant 2.2-fold increase in fusions (p-value <0.05) (Figure 5.11 A). Phleomycin treatment of these cells resulted in a very statistically significant 2.3-fold increase in chromosomal/chromatid fusions (p-value <0.01) (Figure 5.11 A). Expectedly, phleomycin strongly increased the level of chromatid breaks, whereas the effect of ZEB2 overexpression was moderate (approximately 2-fold), and it did not reach statistical significance (Figure 5.11B).

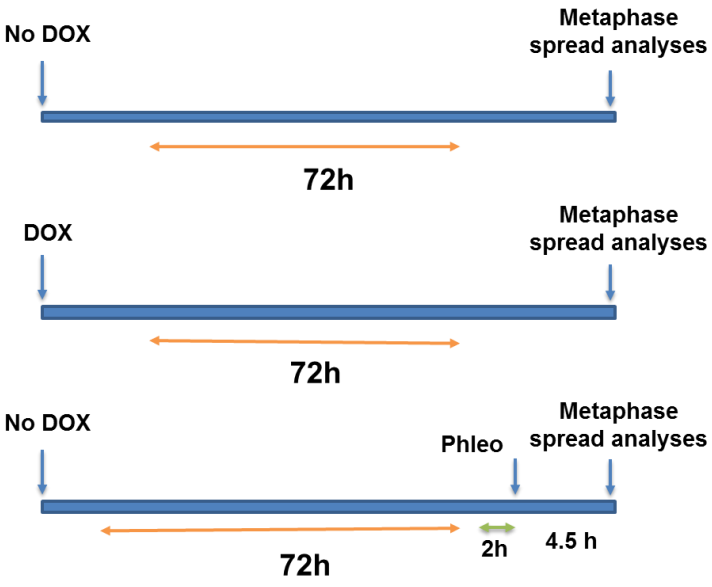


Figure 5.10: A scheme of EMT induction by ZEB2 in A431-ZEB2-cyclin D1 cells karyotyping experiments

showing the duration of the experiment and at which stage the metaphase spread analyses were performed.

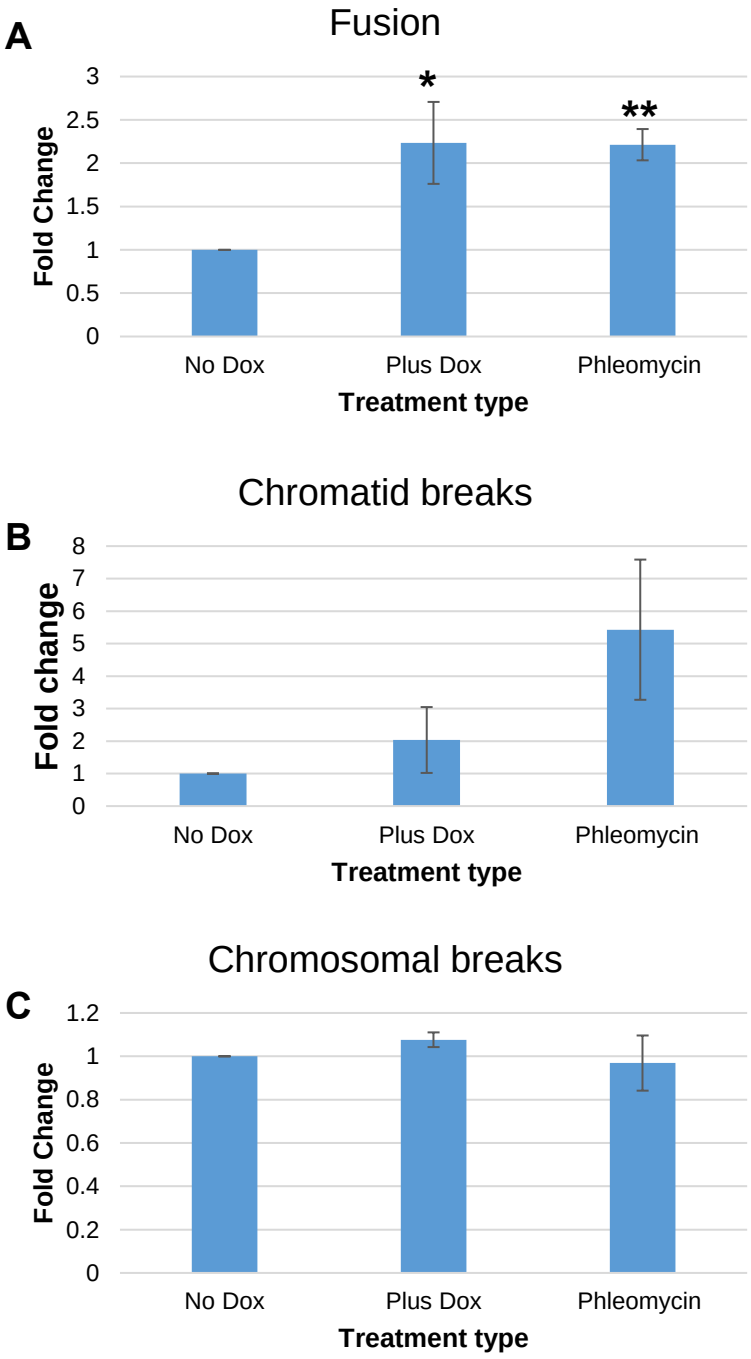


Figure 5.11: Transient EMT induction by ZEB2 in A431-ZEB2-cyclin D1 also results in chromosomal aberrations.

There was an increase in the levels of fusions (A) and chromatid breaks (B) but only a minimal increase in chromosomal breaks (C) as compared to uninduced or Phleomycin treated cells. Dox treatment was 2µg/ml. Phleomycin concentration was 25µg/ml. phleomycin is used as a positive control. ***** p-value <0.05 (0.0377). ****** p-value <0.01 (0.0069). n=3. Number of spreads=35.

5.3.6 Transient ZEB1 overexpression is a factor contributing to aneuploidy

To this end, we have demonstrated that following transient ZEB1 overexpression (i.e. EMT), MCF-7/ZEB1 cells started accumulating DNA damage, as they enter the cell cycle following Dox withdrawal. Furthermore, karyotype analysis has revealed that stable ZEB1 overexpression in proliferating cells, transient ZEB1 overexpression or *LIG1* knockdown in proliferating MCF-7 cells all affect chromosomal integrity and produce aberrations, specially, chromatid breaks and fusions. Here, MCF-7/ZEB1 cells were subjected to either one, two or three rounds of Dox. Next, cells were left to recover for 10 days prior to performing the metaphase spread analyses (see section 2.2.11). This was to determine whether the exposure to multiple rounds of transient ZEB1 overexpression (and EMT) would lead to whole-chromosome gains or losses. MCF7 is an aneuploid cell line with the high average variability in chromosome numbers between individual cells (Yoon et al., 2002). The exposure of the cells to multiple rounds of transient ZEB1 overexpression was found to affect cell distribution according to the chromosome numbers. In MCF-7/ZEB1 cells which never experienced ZEB1 overexpression, the peak of the distribution corresponded to 66 chromosomes per cell, and 50% of cells analysed contained 65-67 chromosomes (Figure 5.12). Remarkably, single round of transient ZEB1 overexpression made this peak less evident (Figure 5.12, 1 Dox), whereas two consecutive rounds of induction/de-induction led to the disappearance of the 65-67 peak. In these cells, 61.3% of the cells contained 60-64 chromosomes. In addition, I detected two-four polyploid cells out of 40 cells karyotyped in a population, which underwent two or three induction/de-induction rounds. This cell population appeared to be much more heterogeneous than the parental cell line, with the average of cells containing fluctuating between 65 to more than 70 chromosomes. Cells with higher chromosome numbers were likely generated by polyploid cells. Overall, my data indicated that ZEB1 overexpression induction/de-induction (i.e. EMT/MET) plasticity is a factor contributing to aneuploidy.

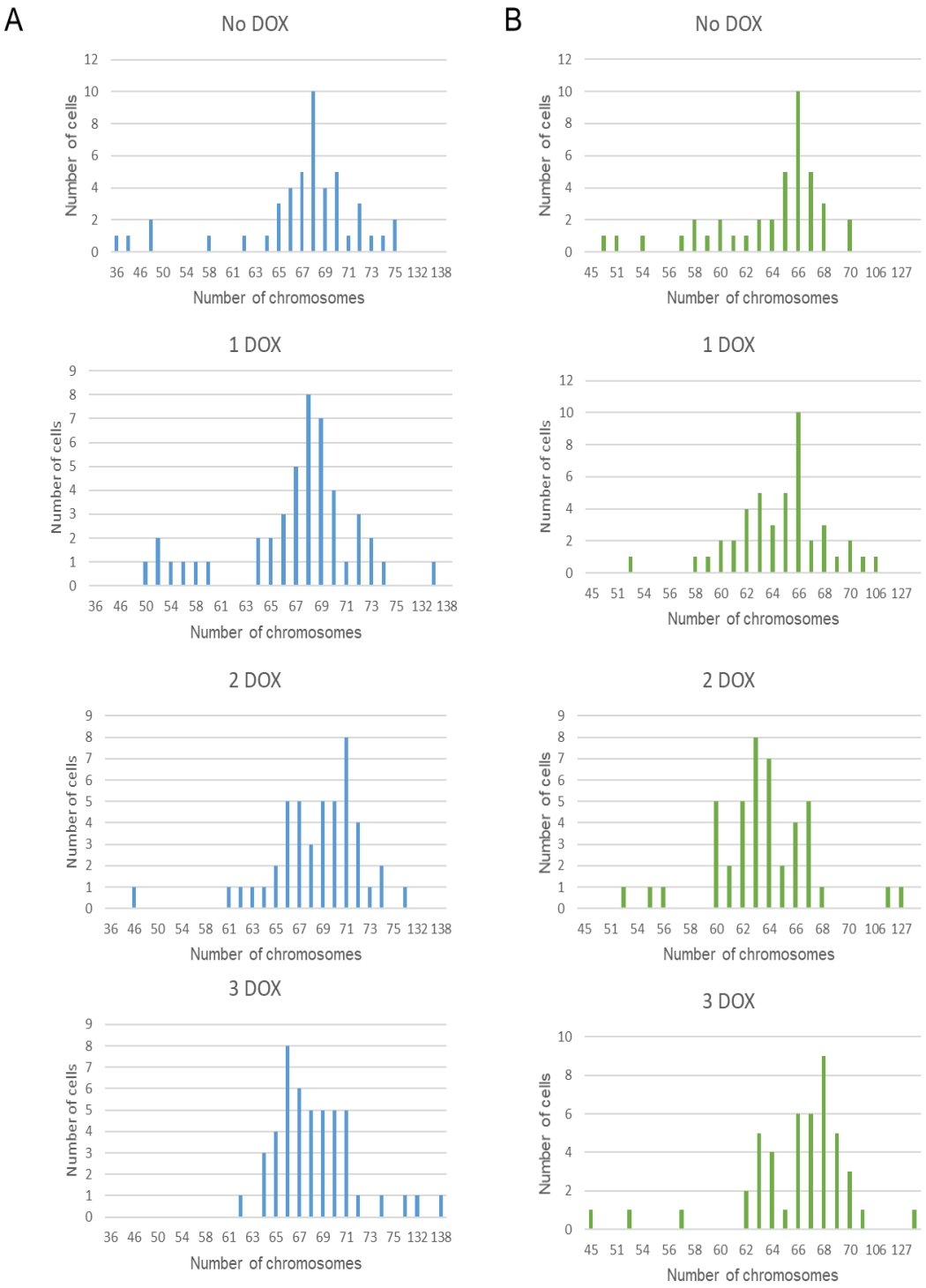


Figure 5.142: Multiple rounds of transient ZEB1 overexpression in MCF-7 cells result in a fluctuation in the average number of chromosomes.

ZEB1 overexpression was induced either once, twice or three times before performing the karyotype analysis. **A** shows the karyotyping results of the first experiment and **B** shows the results for the second experiment. The results indicate that multiple rounds of transient ZEB1 overexpression result in chromosomal instability.

5.4 Discussion

EMT induction has been recently associated with genomic instability in metastatic cancers. Bakhoum et al. (2018) demonstrated that EMT is induced in genetically unstable cancer cells via a mechanism involving spillage of genomic DNA into the cytosol, activation of the DNA-sensing interferon pathway, and noncanonical NF-κB signalling. This work suggests that the defects in chromosomal segregation may cause metastatic spread via EMT pathways. However, EMT may also act upstream of molecular events leading to genomic destabilisation. EMT induction by TGF-β was associated with chromosomal defects in MCF-10A breast epithelial cells, SKBR3 human breast cancer and Comma-D β-geo mouse mammary epithelial cell lines (Comaills et al., 2016). This report documented an increase in the number of binucleated cells, as well as formation of micronuclei in cells undergoing EMT. Those binucleated cells contained multiple centrosomes, thus, causing the missegregation of the chromosomes. EMT induction in proliferating cells was also shown to increase DNA damage (Comaills et al., 2016). This is in line with our findings that γ-H2A.X is present from 8h post Dox withdrawal in MCF-7/ZEB1 cells, hence as cells re-enter the cell cycle (Figure 5.7). We have also shown that the higher levels of DNA damage were detectable at around 8-24h after Dox withdrawal, i.e. at the stage when cells are still deficient in LIG1.

These data are in accord with the studies of cells derived from a patient harbouring a mutation in the *LIG1* gene. These cells accumulate low levels of DNA replication-associated DNA damage (Soza et al., 2009). Moreover, point mutation in mouse *Lig1* that resembled a mutation detected in *LIG1* gene in patients suffering from immunodeficiency and enhanced sun sensitivity, led to the accumulation of SSBs in spleens and thymuses of mutant mice, and occasional bone marrow replication failure. Additionally, spleens of the mutant mice were characterised by high genomic instability and accumulation of replication intermediates. These mice also showed a higher predisposition to rare sporadic epithelial tumours development (Harrison et al., 2002).

We have shown that LIG1 loss in MCF-7 cells resulted in the accumulation of chromosomal aberrations, namely fusions and to a lesser extent breaks (Figure

5.4). This suggests that LIG1 is indispensable for chromosomal integrity in this system. It has been previously proposed that LIG1 could be compensated by LIG3 in some functions in a tissue specific manner (see section 3.4). For instance, it has been shown that LIG3 complexes with XRCC1 to compensate for LIG1 in SSB repair in LIG1-depleted cells (Le Chalony et al., 2012). However, according to our data, although LIG3 was not affected by EMT (Figure 5.7), it could not compensate for LIG1 deficiency in precluding chromosomal defects. This is in line with the finding that LIG3 was unable to compensate for LIG1 in the maintenance of chromosomal integrity by preventing fusions of sister telomeres (Le Chalony et al., 2012). Our data showing enhanced levels of chromosomal/chromatid fusions caused by ZEB1 overexpression or LIG1 KD are reminiscent of the phenotype described by Le Chalony et al. (Figures 5.4, 5.6 and 5.9). Further experiments are required to examine whether the increased level of ZEB1 overexpression-induced fusions was due to telomeres erosion.

It is accepted that EMT in cancer cells is often associated with cell quiescence. In this context, quiescence can be considered as a mechanism protecting cells from genome destabilisation in the course of EMT. Re-entry in the cell cycle after EMT completion is a step in the metastatic cascade, when chromosomal abnormalities are accumulated.

Whole-chromosomal instability is a common form of genetic instability observed in cancer cells. It involves losses and/or gains of whole chromosomes as a consequence of abnormal segregation during mitosis, a phenomenon known as aneuploidy (Thompson and Compton, 2008). Additionally, errors of chromosomal segregation may also result in structural abnormalities and DNA damage (Janssen et al., 2011). We have shown that transient and stable ZEB1 overexpression (i.e. EMT) in MCF-7/ZEB1 cells produced chromosome/chromatid fusions, which could lead to errors in chromosomal segregation during mitosis and generate aneuploid cells. Here we show that subjecting cells to multiple rounds of transient ZEB1 overexpression result in chromosomal instability and aneuploidy incidences, as demonstrated by the variation in the average number of chromosomes in the cells in comparison to uninduced cells (Figure 5.12).

Chapter 6 : General discussion and future directions

EMT-TFs play roles in multiple processes of tumourigenesis such as malignant transformation, therapy resistance, and metastasis. However, only little is known about their role in DDR and genetic instability. Previous studies in our lab demonstrated that EMT induction by ZEB2 overexpression resulted in the reduction of the expression of DDR genes (Dr Gina Tse, Dr Eugene Tulchinsky lab, the University of Leicester). As a result of the vital role of the DDR mechanisms in the maintenance of genomic integrity, this reduction would potentially contribute to genetic instability.

ZEB proteins regulate *LIG1* expression transcriptionally through Rb-E2F

One of the DDR genes that were downregulated by ZEB2 overexpression was the *LIG1* gene, which encodes for the DNA ligase I (LIG1). However, the mechanism by which ZEB2 overexpression regulates LIG1 expression has not yet been unravelled. LIG1 reduction was suggested to be cell cycle dependent, after observing a correlation with cyclin D1 (Dr Gina Tse, Dr Eugene Tulchinsky lab, the University of Leicester). In chapter 3, we demonstrated the same correlation in another model of EMT induction by ZEB proteins overexpression, MCF-7 cells, and we showed that LIG1 expression is regulated transcriptionally in the promoter region. *LIG1* promoter sequence analysis, luciferase assays and CINK4 inhibitor treatment propose E2F dependent regulation. Further experiments, such as chromatin immunoprecipitation should be performed to confirm this interaction.

Thus far, merely one patient with deficiency of LIG1 has been described. This patient suffered from abnormal and delayed growth and development, elevated sun sensitivity, immunodeficiency and lymphoma recurrence. This patient died at the age 19 because of lymphoma (Webster et al., 1992). Studies using cells derived from this patient showed an accumulation of DNA replication-associated DNA damage, both SSBs and DSBs (Soza et al., 2009). Moreover, point mutation in mouse *Lig1* that resembled a mutation detected in *LIG1* gene in patients suffering from immunodeficiency and enhanced sun sensitivity, led to the accumulation of SSBs in spleens and thymuses of mutant mice, and occasional bone marrow replication failure. Additionally, spleens of the mutant mice were characterised by high genomic instability and accumulation of

replication intermediates. These mice also showed a higher predisposition to rare sporadic epithelial tumours development (Harrison et al., 2002).

LIG1 loss in MCF-7 results in chromosomal aberrations

LIG1 is involved in DNA repair and replication. Therefore, reduction in its expression in the G1 phase would result in the accumulation of the SSBs that will be converted to DSBs in the S phase when DNA replication occurs (Soza et al., 2009). As the HR process is defective as well upon EMT induction by ZEB2 (Dr Gina Tse, Dr. Eugene Tulchinsky Laboratory, the University of Leicester), those breaks will be repaired by the less accurate NHEJ leading to chromosomal instability and an increase in genomic instability. We have shown in chapter 5 that LIG1 loss in MCF-7 cells resulted in the accumulation of chromosomal aberrations, namely chromatid/chromosomal fusions and to a lesser extent chromatid breaks. This suggests that LIG1 is indispensable for chromosomal integrity in this system. It was previously proposed that LIG1 could be compensated by LIG3 in some functions in a tissue specific manner (see section 3.4). However, according to our data, although LIG3 was not affected by ZEB1 overexpression, it could not compensate for LIG1 deficiency in precluding chromosomal defects. This is in line with the finding that LIG3 was unable to compensate for LIG1 in the maintenance of chromosomal integrity by preventing fusions of sister telomeres (Le Chalony et al., 2012). Our data showing enhanced levels of chromosomal/chromatid fusions caused by ZEB1 overexpression or LIG1 KD are reminiscent of the phenotype described by (Le Chalony et al., 2012). Further experiments are required to examine whether the increased level of ZEB1 overexpression-induced fusions was due to telomeres erosion.

ZEB1 overexpression in MCF-7 cells results in cell cycle arrest at G1

In chapter 4, we have shown that ZEB1 overexpression in MCF-7 cells arrested the cells at G1. Western blotting and RNAi experiments demonstrated that this arrest is a result of p27Kip1 expression upregulation. Luciferase assays proposed a direct activation of the *CDKN1B* promoter by ZEB1. Further experiments are required to confirm this mechanism of regulation.

Although p27Kip1 is a potent tumour suppressor, its genetic mutations are very rare (Alkarain and Slingerland, 2004; He et al., 2012). However, in many human

cancers, p27Kip1 protein level is usually lost or decreased. This loss or reduction was shown to associate with decreased patient survival and poor prognosis in many cancer types including carcinomas of breast, gastric, prostate, colon and lung (Lacy et al., 2005; Slingerland and Pagano, 2000). This reduction in p27Kip1 is usually a consequence of an enhanced protein degradation. For instance, Gstaiger et al. (2001) shown that p27Kip1 expression was reduced or absent in tumours with high expression of the ubiquitin ligase Skp2, and vice versa. Similar inverse relationship between p27Kip1 and Skp2 expression levels were demonstrated in subsets of many cancers, including colorectal, prostatic, small cell lung, gastric, oral carcinomas, and lymphomas (Chiarle et al., 2000; Hershko et al., 2001; Kudo et al., 2001; Masuda et al., 2002; Yokoi et al., 2002). Although mutations of *CDKN1B* gene are very rare in cancers, *CDKN1B* mutations has been identified in breast cancer (Stephens et al., 2012; Tigli et al., 2005).

Here we suggest a tumour suppressive and protective role of p27Kip1 after ZEB1 overexpression in this system by inducing quiescence. Our experiments demonstrated that loss of p27Kip1, by RNAi, merely uncoupled quiescence from other effects induced by ZEB1 overexpression. Cells with p27Kip1 depletion were found to be LIG1 deficient. Therefore, without the p27Kip1 expression, cells would be proliferating while lacking LIG1 resulting in chromosomal instability. This is in line with the recently suggested association between p27Kip1 loss and the elevation of genomic instability and the induction of radioresistance in luminal breast cancer cells (Berton et al., 2017). In chapter 5, we have demonstrated this by performing karyotype analysis in a model of a stable ZEB1 overexpression that lacks p27Kip1 expression and found an increase in chromosomal aberrations and namely chromatid/chromosomal fusions. This is in line with the previously suggested generation of genomic instability due to continued proliferation of cells undergoing EMT (Comaills et al., 2016). Therefore, here we suggest that the source of genomic instability in this system is at the step of the metastatic cascade at which the cells re-enter the cell cycle (Figure 6.1).

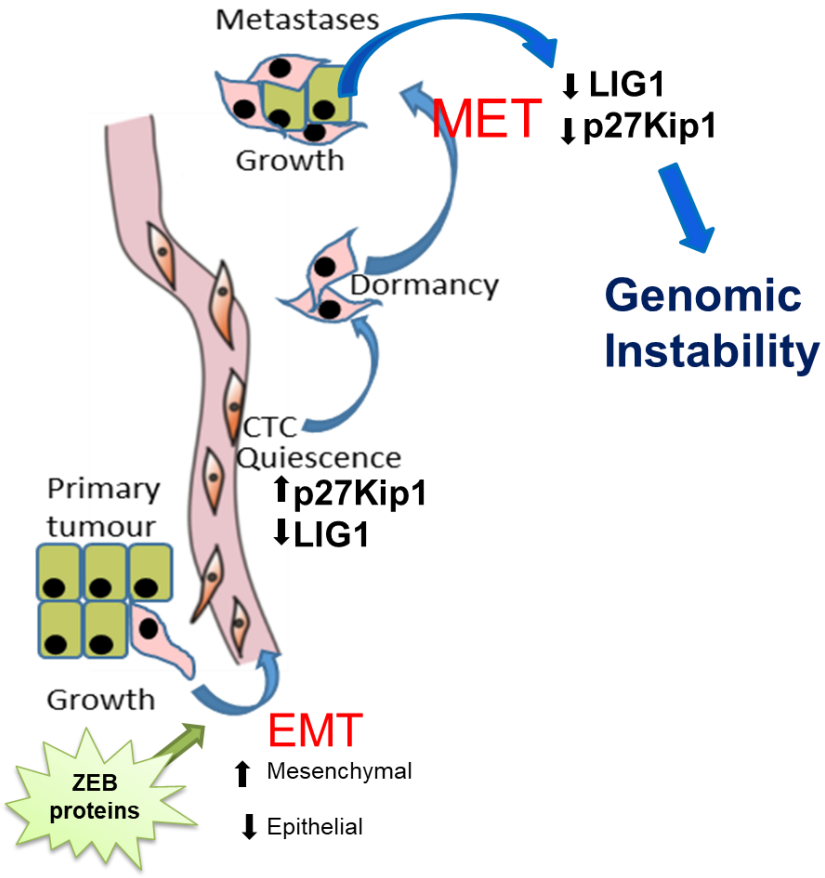


Figure 6.1: A schematic representation of how exit from dormancy could be a source of genomic instability.

After the induction of ZEB proteins expression in the primary tumour, EMT induction leads to the formation of a pool of dormant circulating tumour cells, that have an upregulation of p27Kip1 expression and a reduction in the expression of LIG1. These cells then intravasate into the blood or lymphatic systems, spread, extravasate and form metastases at distant organs. Formation of the metastases is often associated with the mesenchymal epithelial transition (MET) of tumour cells. We found that cell cycle re-entry might be mutagenic, associated with the reduced *LIG1* expression, and enhanced level of chromatid or chromosomal fusions. We propose that transient ZEB overexpression contribute to genomic instability in cancer. Image is taken from Dr Eugene Tulchinsky, University of Leicester.

EMT role in chromosomal instability and genomic integrity has an element of controversy. On the one hand, Zhang et al. (2014a); Zhang et al. (2014b) have suggested that EMT induction serves a protective role on genomic integrity through the induction of radioresistance. Radioresistance was generated through enhancement of HR-mediated repair of DSB, which improves cell survival. Additionally, another study has demonstrated that oncogenic transduction in non-transformed mammary stem cells, by cyclin E or Ras, resulted in the initiation of a ZEB1- methionine sulfoxide reductase (MSRB3) axis, which enables them to resist oncogene-induced DNA damage, maintain genome integrity and protects them from ROS-induced apoptosis (Morel et al., 2017).

On the other hand, EMT induction was shown to induce genetic instability. Comaills et al. (2016) proposed that EMT induction by TGF-β and SNAIL resulted in the generation of micronuclei and binucleation, which led to chromosomal missegregation during mitosis culminating in aneuploidy. Furthermore, a recent study has shown that the spillage of genomic DNA into the cytosol as a result of chromosomal segregation errors in genetically unstable cancer cells results in the activation of the DNA-sensing interferon pathway, and noncanonical NF-κB signalling. This eventually results in EMT activation and inflammation and is a driver of metastatic spread (Bakhoum et al., 2018).

So, Bakhoum et al. (2018) suggest that genetic instability lies upstream of EMT while our data and Comaills et al. (2016) demonstrated a role of EMT upstream of genetic instability. This could be proposing a positive feedback loop between EMT and genetic instability, where genetic instability induces EMT, which in turn enforces further instability. I propose that this self-enforcing double positive loop is a driving force of tumourigenesis.

Are these effects on LIG1 and p27Kip1 a result of ZEB proteins overexpression or EMT induction?

Answering this question is dependent on how EMT is defined. As if EMT was only defined using the classical view, which is the loss of epithelial markers and the concomitant gain of mesenchymal ones, then any other feature of induced after EMT-TFs induction would be an EMT independent feature (Brabletz et al., 2018). Previously, EMT-TFs were defined merely as inducers of EMT and their

implication in cells invasion and dissemination. However, the new insight defines them as versatile transcription factors that are implicated in pleotropic functions (Figure 6.2) (Goossens et al., 2017). As discussed earlier, EMT definition is not simply a complete change from epithelial to mesenchymal phenotype. Instead, based on the complexity and plasticity of the process, a cell may merely undergo partial transition with this stage being the last differentiation point for the cell (Brabletz et al., 2018; Nieto et al., 2016). In fact, this can also explain the contribution of EMT-TFs in the malignant progression of several non-carcinomatous tumours. As merely traits like stemness, CSC generation, therapy resistance and invasiveness were observed in these tumours, rather than classic features of the EMT programmes (Brabletz et al., 2018).

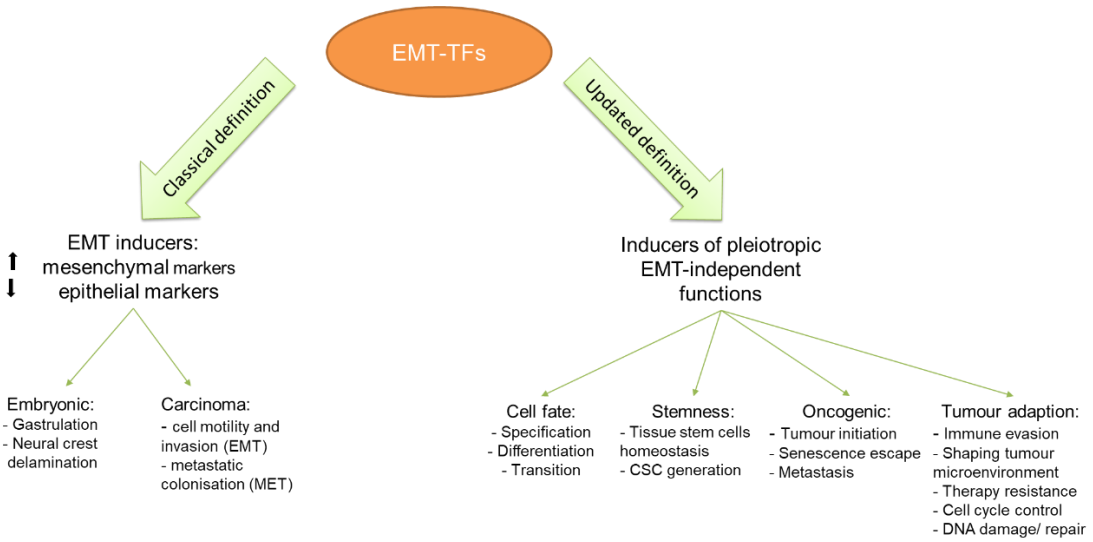


Figure 6.2: a scheme of the re-evaluated insights of EMT-TFs functions in normal conditions and in disease.

The EMT-TFs of the SNAIL, ZEB and TWIST family were defined merely as EMT inducers for nearly two decades. However, based on the recent research of these TFs, this scope has been extended and re-evaluated. The updated definition suggests that these EMT-TFs also function to induce various functions and not just limited to the classical EMT features (i.e. invasion and dissemination) both in physiological and pathological conditions. The figure is adapted from (Goossens et al, 2017).

EMT-TFs of the ZEB, SNAIL and TWIST families were recently shown to be involved in various pivotal functions of cancer other than invasion and dissemination (Puisieux et al., 2014). They were shown to have oncogenic

functions that facilitate cancer initiation and progression. Examples of these functions include controlling the plasticity of stem cells and cancer stem cells (Brabletz, 2012; Guo et al., 2012; Mani et al., 2008; Morel et al., 2008; Schmidt et al., 2015), tumour initiation and malignant transformation (De Craene and Berx, 2013; De Craene et al., 2013; Li et al., 2017a; McCormack et al., 2010), therapy resistance (Cojoc et al., 2015; Meidhof et al., 2015; Zhang et al., 2014b), immune evasion and cancer micro-environment shaping (Lyons et al., 2008; Omilusik et al., 2015; van Helden et al., 2015) and cell cycle control (Mejlvang et al., 2007; Sayan et al., 2009; Vega et al., 2004).

Here we provide a potential role for the ZEB family of EMT-TFs in DNA damage response and cell cycle control. We showed that ZEB1/2 proteins overexpression resulted in the downregulation of LIG1 expression in Rb-E2F-dependent manner. We also showed that ZEB overexpression arrests cells at G1 of the cell cycle in a mechanism that involves p27Kip1 upregulation. This ultimately results in the accumulation of DNA damage, chromosomal aberrations and aneuploidy. Our findings provide new insights to the contribution of the ZEB family of EMT-TFs to genomic instability in carcinomas.

Limitations and future directions

Finally, several important limitations need to be considered. One of the limitations is the absence of controls for EMT induction and effects in the cell lines used. Although these controls was not analysed in depth for this study, both cell lines were extensively analysed for EMT induction features and markers previously in our lab (for MCF-7 cells, Dr Youssef.Alghamdi, Dr Eugene Tulchinsky Lab, University of Leicester, and for A431-ZEB2-cyclin D1, (Mejlvang et al., 2007)). Another limitation is the absence of ZEB proteins background level in the parental MCF-7 cell line (i.e. before the introduction of ZEB1/ZEB2 GFP constructs). Additionally, the parental cell lines were not tested for Dox treatment effects, to confirm that the observed phenotype is specific to ZEB proteins overexpression.

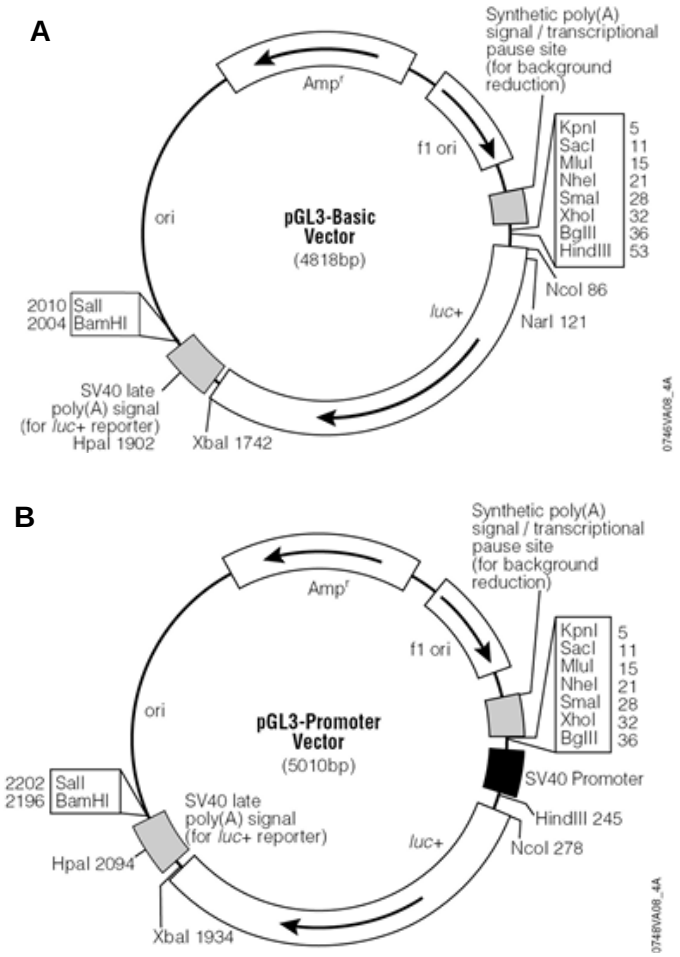
Our findings provided several insights for future research. For instance, we have proposed an E2F dependent regulation of *LIG1* promoter, to confirm this interaction, further experiments, such as chromatin immunoprecipitation, and site directed mutagenesis, should be performed. Additionally, p27Kip1 loss in ZEB1

overexpressing MCF-7 cells was found to release the cell cycle arrest but without alleviating ZEB1 effect on the expression of LIG1, cyclin D1 and Rb. Further experiments should be performed to confirm p27Kip1 loss on other features of ZEB1 overexpression, such as invasiveness.

Furthermore, ZEB1 overexpression in these cells was found to increase the transcription level of the *CDKN1B* gene as well as reduce the expression of a potential targeting miRNA (miR200c). Further investigations then suggested a possible regulation by miRNAs and at the promoter level. More experiments should be done to confirm this regulation. For instance, for the miRNA responsiveness, other members of the miR200 family should be tested. Additionally, miRNA mimics can be used to see whether a similar effect could be achieved. Moreover, site directed mutagenesis could be used to modify the miRNA binding site to see whether the inhibition can be abrogated.

ɣ-H2A.X staining suggested an increase of DNA damage as cells re-enter the cell cycle, after Dox withdrawal and ZEB1 expression starts to decrease. To confirm that this effect is due to ZEB1 reduction and not due to Dox treatment, parental cell line should be treated with Dox to see whether the same DNA can be observed. In addition, the karyotype of parental MCF-7 cell line treated with Dox, once and multiple times, should be analysed to confirm that the observed chromosomal aberrations are a result of LIG1 knockdown, transient ZEB1 overexpression and stable ZEB1 overexpression coupled with *CDKN1B* knockdown.

Appendices



Appendix 1: Maps of the pGL3-basic (A) and pGL3-promoter (B) luciferase reporter plasmids. The maps show the restriction sites of several enzymes, position of the SV40 promoter, type of antibiotic resistance and the size of the plasmids.

Score	Expect	Identities	Gaps	Strand	Frame
1945 bits(1053)	0.0()	1079/1097(98%)	4/1097(0%)	Plus/Plus	

Features:

```
Query  1399  AAAGGTCCATTTTCCAGGGCAGATTTAAATTTGGTTAAGATCTGATTTCTTGGCCAGGTG  1458
             |||||||||| ||||||| |||||||||||| ||  ||||   ||||| ||||||||||
Sbjct  1     AAAGGTCCATNTTCCAGGNCAGATTTAAATTNGGNNAAGANNNGATTTNTTGGCCAGGTG  60

Query  1459  CAGTGGCTCACGCCTGTACTCCCAGCAGTTTGGGAGGCCGAAGCGGGTGGATCACTTGAG  1518
             ||  |||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Sbjct  61    CANNGGCTCACGCCTGTATTCCCAGCAGTTTGGGAGGCCGAAGCGGGTGGATCACTTGAG  120

Query  1519  GTCGGGAGTTGGAGACCAGCCTGGCCAACATGGTGAAACCCCATCCCTACCAAAATTATA  1578
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Sbjct  121   GTCGGGAGTTGGAGACCAGCCTGGCCAACATGGTGAAACCCCATCCCTACCAAAATTACA  180

Query  1579  AAAATTAGCTGGGTATGGTGGCATGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCA  1638
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  181   AAAATTAGCTGGGTATGGTGGCATGCGCCTGTAGTCCCAGCTACTCAGGAGGCTGAGGCA  240

Query  1639  TGAGAATCCCTTGAAACTGGGAGGTGGAGGTTGCAGTGAGCCGAGATTGTGCCACTGCAC  1698
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  241   TGAGAATCCCTTGAAACTGGGAGGTGGAGGTTGCAGTGAGCCGAGATTGTGCCACTGCAC  300

Query  1699  TCTAGCCTGGGCGACAGAGGGAGACTGTGTCTCaaaaaaaaaaGATTTGATTTCTTATTG  1758
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  301   TCTAGCCTGGGCGACAGAGGGAGACTGTGTCTCAAAAAAAAAAGATTTGATTTCTTATTG  360

Query  1759  ACAATTTTTAGTACAAATTTTATTGTAAACTACACTTTTAGCCACTGTCAGTGTCCAGTG  1818
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Sbjct  361   ACAATTTTTAGTACAAATTTTATTGTAAACTACACTNTTAGCCACTGTCAGTGTCCAGTG  420

Query  1819  AGGGGTGGCGGGGGTGGCCTCTTTCTACTAAGGCAGATTTTTCCCAAGTTGGTTTTTGTA  1878
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  421   AGGGGTGGCGGGGGTGGCCTCTTTCTACTAAGGCAGATTTTTCCCAAGTTGGTTTTTGTA  480

Query  1879  GTTTGTATTGTGACCCTAGAGGTAAGAttttttttttaactttgaagatttttatttttg  1938
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  481   GTTTGTATTGTGACCCTAGAGGTAAGATTTTTTTTTTAACTTTGAAGATTTTTATTTTTG  540

Query  1939  tctataacaggctgtcatttgcatttttaacccttttCAGATTCTATTTTGGGGTGAATAA  1998
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  541   TCTATAACAGGCTGTCATTTGCATTTTTAACCCTTTCAGATTCTATTTTGGGGTGAATAA  600

Query  1999  CACTAAAGGAATCATTGGTAAGGAAAGACACACCATTTATGTCTCGGCTACATAGGCAAA  2058
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601   CACTAAAGGAATCATTGGTAAGGAAAGACACACCATTTATGTCTCGGCTACATAGGCAAA  660

Query  2059  C--GGGCTGGACCCCCTCAAGGTCGCTGCCTTGCTTGGGGTTGACAAGGAGCAGGGTCTT  2116
             |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661   CAGGGGCTGGACCCCCTCAAGGTCGCTGCCTTGCTTGGGGTTGACAAGGAGCAGGGTCTT  720

Query  2117  ACGTGGCCCACTGCTTTACAGGCACATTGGATTGGCACAAGATAGAACTGTAAGACCTGC  2176
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  721   ACGTGGCCCACTGCTTTACAGGCACATTGGATTGGCACAAGATAGAACTGTAAGACCTGC  780

Query  2177  ACTTTGGGAGGCCAAGGCGGGTGGCTCACCTGAGGTCAGTAGTTCAAGACCAGCCTGGCC  2236
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  781   ACTTTGGGAGGCCAAGGCGGGTGGCTCACCTGAGGTCAGTAGTTCAAGACCAGCCTGGCC  840

Query  2237  AACATGGTGAAATCCCATCTCTACTAAAAATACaaaaaaaTAGCTGGATGTGGTGGTGGG  2296
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  841   AACATGGTGAAATCCCATCTCTACTAAAAATACAAAAAAATAGCTGGATGTGGTGGTGGG  900

Query  2297  CGCCTGTCATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCATGAACCCAGGAGGC  2356
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  901   CGCCTGTCATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCATGAACCCAGGAGGC  960

Query  2357  AGAGGTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCTGGGCGGTAAGAACGAAA  2416
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  961   AGAGGTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCTGGGCGGTAAGAACGAAA  1020

Query  2417  CTCTGTCTCaaaaaaaaGAACAGATGACAGATGAGGGCAGACTGTCTCCAGAGTCAAAAT  2476
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1021  CTCTGTCTCAAAAAAAAGAACAGATGACAGATGAGGGCAGACTGTCTCCAGAGTCAAAAT  1080

Query  2477  GCCCTTGTAGGTTCAAT  2493
             ||||| |||||| ||||
Sbjct  1081  GCCCT-GTAGGT-CAAT  1095
```

Appendix 2: pGL3-Promoter-*LIG1* enh-2532 plasmid confirmation.

Score	Expect	Identities	Gaps	Strand	Frame
4676 bits(2532)	0.0()	2532/2532(100%)	0/2532(0%)	Plus/Plus	

Features:

```
Query  3715   GTGAGCCCCCAGAAGGAGAGAAgtgtgtgtgtgtgtgtgtgtgtgtgtgaagccaaga  3774
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21433  GTGAGCCCCCAGAAGGAGAGAAGTGTGTGTGTGTGTGTGTGTGTGTGAAGCCAAGA  21492

Query  3775   ggaaggaggaggaggaaggtgagcccccagaaggagaggtgtgtgtgtgtgtgtgtgt  3834
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21493  GGAAGGAGGAGGAGGAAGGTGAGCCCCCAGAAGGAGAGGTGTGTGTGTGTGTGTGTGT  21552

Query  3835   gtgtgtgtgtgtgaagccaagaggaaggaggaggaggaaggtgagcccccagaaggagag  3894
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21553  GTGTGTGTGTGTGAAGCCAAGAGGAAGGAGGAGGAGGAAGGTGAGCCCCCAGAAGGAGAG  21612

Query  3895   gtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtTTTGGA  3954
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21613  GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTTGGA  21672

Query  3955   TCTGCAAGGGCAGGGGCCTGAGAGCTCAGGATGTTTGGACCCTGCCGATTTGATGCCTGT  4014
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21673  TCTGCAAGGGCAGGGGCCTGAGAGCTCAGGATGTTTGGACCCTGCCGATTTGATGCCTGT  21732

Query  4015   GTCCAGGCATCCAGGAATTCCACCTACCAATATGGGAACTTTTGTTCTGAAAACATTGTG  4074
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21733  GTCCAGGCATCCAGGAATTCCACCTACCAATATGGGAACTTTTGTTCTGAAAACATTGTG  21792

Query  4075   GGATATATTTGTTTGGACATCTTAATTTAGTAATAATTCAGGCACAGTTTCCCatttta  4134
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21793  GGATATATTTGTTTGGACATCTTAATTTAGTAATAATTCAGGCACAGTTTCCCATTTTTA  21852

Query  4135   tttgctatttttatatttatttaatggaaactatttttcctctaaatgatttccttttCC  4194
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21853  TTTGCTATTTTTATATTTATTTAATGGAAACTATTTTTCCTCTAAATGATTTCCTTTTCC  21912

Query  4195   TTGTAAATTTTACCATAAAGTTGCAGTTCCCAAACATTCTTAGTTCACAGCACCCTTGAT  4254
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21913  TTGTAAATTTTACCATAAAGTTGCAGTTCCCAAACATTCTTAGTTCACAGCACCCTTGAT  21972

Query  4255   GTCTTAGTAAttttttATGATGCTCCTAGGCCAATACCTTATCAGTTCTATTAAGTAGT  4314
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  21973  GTCTTAGTAATTTTTTTATGATGCTCCTAGGCCAATACCTTATCAGTTCTATTAAGTAGT  22032

Query  4315   TATGTCCAAACAACTATTAGATACTATACATACATTGGaaaaaaaGTAATATTTGTGCtt  4374
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22033  TATGTCCAAACAACTATTAGATACTATACATACATTGGAAAAAAAGTAATATTTGTGCTT  22092

Query  4375   ttttttttttttttttttGAGACAGTCTCACTGTGTCGCTCAGGCTACAGTGCAGTGGC  4434
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22093  TTTTTTTTTTTTTTTTTTGAGACAGTCTCACTGTGTCGCTCAGGCTACAGTGCAGTGGC  22152

Query  4435   ACAGTTTGGGCTCACTGCAGTCTCCACCTCCCAGGCTCAAGCGATTCTCCCACCTCAGCT  4494
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22153  ACAGTTTGGGCTCACTGCAGTCTCCACCTCCCAGGCTCAAGCGATTCTCCCACCTCAGCT  22212

Query  4495   TCCCGAGTAGCTGGGATTACAGACACACCACCACGCCCAGCTAATTTTTGTATTTTTAGT  4554
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22213  TCCCGAGTAGCTGGGATTACAGACACACCACCACGCCCAGCTAATTTTTGTATTTTTAGT  22272

Query  4555   AGAGACAGGGTTTCACTATGTTGACCAAGCTGGTCTCGAACTCCTGAGCTCACATGATCT  4614
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22273  AGAGACAGGGTTTCACTATGTTGACCAAGCTGGTCTCGAACTCCTGAGCTCACATGATCT  22332

Query  4615   GCCCACCTCGGCCTCACAAAGCCCTGGGATTACAGGCGTGAGCCACTGCGCCCAGCCTGT  4674
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22333  GCCCACCTCGGCCTCACAAAGCCCTGGGATTACAGGCGTGAGCCACTGCGCCCAGCCTGT  22392

Query  4675   ATTTTGTTCTTAAATAACCATAATTACTTGTGGGAAATGCCTTGGATACTCTGTAACATC  4734
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22393  ATTTTGTTCTTAAATAACCATAATTACTTGTGGGAAATGCCTTGGATACTCTGTAACATC  22452

Query  4735   TTAAACAGTGGAATTCAATTGAAAACTGCCAACGTCATTTCCTGTTCTATGTTGATTTTT  4794
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22453  TTAAACAGTGGAATTCAATTGAAAACTGCCAACGTCATTTCCTGTTCTATGTTGATTTTT  22512

Query  4795   GTACAGTACTTGATTTTTATAGCAGCAACCTATGAAAACCCAGCTTTGCAGAGATGCAAT  4854
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22513  GTACAGTACTTGATTTTTATAGCAGCAACCTATGAAAACCCAGCTTTGCAGAGATGCAAT  22572

Query  4855   GTCATTGCAAGGAATGTGATTTAATACTGAAAACTGAAATACCTCGAACTAGTAGTTCAC  4914
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22573  GTCATTGCAAGGAATGTGATTTAATACTGAAAACTGAAATACCTCGAACTAGTAGTTCAC  22632
```

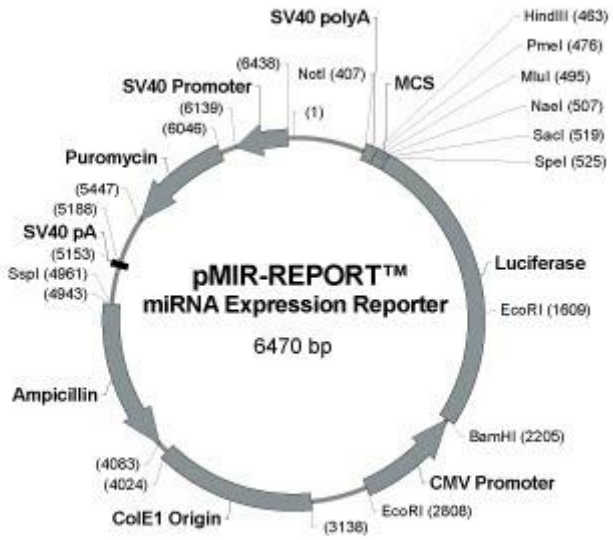
Appendix 3: pGL3- *LIG1* pr-1727-*LIG1* enh-2532 plasmid confirmation.

```
Query  4915   ACACTGGCCAGCAGGTGTCGCTGGGTTTCCTTTGAAAAATCTAAAATATCTTGCAGCGCC  4974
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22633  ACACTGGCCAGCAGGTGTCGCTGGGTTTCCTTTGAAAAATCTAAAATATCTTGCAGCGCC  22692

Query  4975   CAGTGAGTTCCCTGCATGTTGGTGAACCCACCGAAAACAGCCATAAAGATCTGATCGATT  5034
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22693  CAGTGAGTTCCCTGCATGTTGGTGAACCCACCGAAAACAGCCATAAAGATCTGATCGATT  22752

Query  5035   GATCATCTGGGGCCTGCATGAGATTACACCAATTGCAGCCTTTAACGTCATTTGGTTTTT  5094
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22753  GATCATCTGGGGCCTGCATGAGATTACACCAATTGCAGCCTTTAACGTCATTTGGTTTTT  22812

Query  5095   TGTTTCTGCAAACAAATAAAAGGTCCATTTTCCAGGGCAGATTTAAATTTGGTTAAGATC  5154
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22813  TGTTTCTGCAAACAAATAAAAGGTCCATTTTCCAGGGCAGATTTAAATTTGGTTAAGATC  22872

Query  5155   TGATTTCTTGGCCAGGTGCAGTGGCTCACGCCTGTACTCCCAGCAGTTTGGGAGGCCGAA  5214
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22873  TGATTTCTTGGCCAGGTGCAGTGGCTCACGCCTGTACTCCCAGCAGTTTGGGAGGCCGAA  22932

Query  5215   GCGGGTGGATCACTTGAGGTCGGGAGTTGGAGACCAGCCTGGCCAACATGGTGAAACCCC  5274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22933  GCGGGTGGATCACTTGAGGTCGGGAGTTGGAGACCAGCCTGGCCAACATGGTGAAACCCC  22992

Query  5275   ATCCCTACCAAAATTATAAAAATTAGCTGGGTATGGTGGCATGCGCCTGTAGTCCCAGCT  5334
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  22993  ATCCCTACCAAAATTATAAAAATTAGCTGGGTATGGTGGCATGCGCCTGTAGTCCCAGCT  23052

Query  5335   ACTCAGGAGGCTGAGGCATGAGAATCCCTTGAAACTGGGAGGTGGAGGTTGCAGTGAGCC  5394
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23053  ACTCAGGAGGCTGAGGCATGAGAATCCCTTGAAACTGGGAGGTGGAGGTTGCAGTGAGCC  23112

Query  5395   GAGATTGTGCCACTGCACTCTAGCCTGGGCGACAGAGGGAGACTGTGTCTCaaaaaaaaa  5454
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23113  GAGATTGTGCCACTGCACTCTAGCCTGGGCGACAGAGGGAGACTGTGTCTCAAAAAAAAA  23172

Query  5455   aGATTTGATTTCTTATTGACAATTTTTAGTACAAATTTTATTGTAAACTACACTTTTAGC  5514
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23173  AGATTTGATTTCTTATTGACAATTTTTAGTACAAATTTTATTGTAAACTACACTTTTAGC  23232

Query  5515   CACTGTCAGTGTCCAGTGAGGGGTGGCGGGGGTGGCCTCTTTCTACTAAGGCAGATTTTT  5574
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23233  CACTGTCAGTGTCCAGTGAGGGGTGGCGGGGGTGGCCTCTTTCTACTAAGGCAGATTTTT  23292

Query  5575   CCCAAGTTGGTTTTTGTAGTTTGTATTGTGACCCTAGAGGTAAGAttttttttttaactt  5634
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23293  CCCAAGTTGGTTTTTGTAGTTTGTATTGTGACCCTAGAGGTAAGATTTTTTTTTTAACTT  23352

Query  5635   tgaagatttttatttttgtctataacaggctgtcatttgcatttttaacccttCAGATT  5694
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23353  TGAAGATTTTTATTTTTGTCTATAACAGGCTGTCATTTGCATTTTTAACCCTTTCAGATT  23412

Query  5695   CTATTTTGGGGTGAATAACACTAAAGGAATCATTGGTAAGGAAAGACACACCATTTATGT  5754
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23413  CTATTTTGGGGTGAATAACACTAAAGGAATCATTGGTAAGGAAAGACACACCATTTATGT  23472

Query  5755   CTCGGCTACATAGGCAAACGGGCTGGACCCCCTCAAGGTCGCTGCCTTGCTTGGGGTTGA  5814
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23473  CTCGGCTACATAGGCAAACGGGCTGGACCCCCTCAAGGTCGCTGCCTTGCTTGGGGTTGA  23532

Query  5815   CAAGGAGCAGGGTCTTACGTGGCCCACTGCTTTACAGGCACATTGGATTGGCACAAGATA  5874
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23533  CAAGGAGCAGGGTCTTACGTGGCCCACTGCTTTACAGGCACATTGGATTGGCACAAGATA  23592

Query  5875   GAACTGTAAGACCTGCACTTTGGGAGGCCAAGGCGGGTGGCTCACCTGAGGTCAGTAGTT  5934
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23593  GAACTGTAAGACCTGCACTTTGGGAGGCCAAGGCGGGTGGCTCACCTGAGGTCAGTAGTT  23652

Query  5935   CAAGACCAGCCTGGCCAACATGGTGAAATCCCATCTCTACTAAAAATACaaaaaaaTAGC  5994
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23653  CAAGACCAGCCTGGCCAACATGGTGAAATCCCATCTCTACTAAAAATACAAAAAAATAGC  23712

Query  5995   TGGATGTGGTGGTGGGCGCCTGTCATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCG  6054
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23713  TGGATGTGGTGGTGGGCGCCTGTCATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCG  23772

Query  6055   CATGAACCCAGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCTG  6114
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23773  CATGAACCCAGGAGGCAGAGGTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCTG  23832

Query  6115   GGCGGTAAGAACGAAACTCTGTCTCaaaaaaaaGAACAGATGACAGATGAGGGCAGACTG  6174
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23833  GGCGGTAAGAACGAAACTCTGTCTCAAAAAAAAGAACAGATGACAGATGAGGGCAGACTG  23892
```

Appendix 3: pGL3- *LIG1* pr-1727-*LIG1* enh-2532 plasmid confirmation.

```
Query  6175   TCTCCAGAGTCAAAATGCCCTTGTAGGTTCAATGGGAttttttGCCATCTGACCGTTCT  6234
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  23893  TCTCCAGAGTCAAAATGCCCTTGTAGGTTCAATGGGATTTTTTTGCCATCTGACCGTTCT  23952

Query  6235   GTCTATCGTCAG  6246
              ||||||||||||
Sbjct  23953  GTCTATCGTCAG  23964
```

Appendix 3: pGL3- *LIG1* pr-1727-*LIG1* enh-2532 plasmid confirmation.



Appendix 4: The map of the pMIR-REPORT miRNA expression luciferase reporter used to generate the *CDKN1B* 3'UTR plasmids. The map shows the sites for restriction enzymes, antibiotic resistance and the plasmid size.

Score	Expect	Identities	Gaps	Strand	Frame
1151 bits(623)	0.0()	642/653(98%)	4/653(0%)	Plus/Plus	

Features:

```
Query  3    ATGTTTCCTTGTTTATCAGNATACATCACTGCTTGATGAAGCAAGGAAGATNNNNNNATG  62
            |||||||||||||||||| ||||||||||||||||||||||||||||||      |||
Sbjct  29   ATGTTTCCTTGTTTATCAG-ATACATCACTGCTTGATGAAGCAAGGAAGAT--ATACATG  85

Query  63   AAAATTTTAAAAATACATATCGCTGACTTCATGGAATGGACATCCTGTATAAGCACTGAA  122
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  86   AAAATTTTAAAAATACATATCGCTGACTTCATGGAATGGACATCCTGTATAAGCACTGAA  145

Query  123  AAA-CAACAACACAATAACACTAAAATTTTAGGCACTCTTAAATGATCTGCCTCTAAAAG  181
            ||| ||||| | ||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  146  AAANCAACACCCCAATAACACTAAAATTTTAGGCACTCTTAAATGATCTGCCTCTAAAAG  205

Query  182  CGTTGGATGTAGCATTATGCAATTAGGTTTTTCCTTATTTGCTTCATTGTACTACCTGTG  241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  206  CGTTGGATGTAGCATTATGCAATTAGGTTTTTCCTTATTTGCTTCATTGTACTACCTGTG  265

Query  242  TATATAGTTTTTACCTTTTATGTAGCACATAAACTTTGGGGAAGGGAGGGCAGGGTGGGG  301
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  266  TATATAGTTTTTACCTTTTATGTAGCACATAAACTTTGGGGAAGGGAGGGCAGGGTGGGG  325

Query  302  CTGAGGAACTGACGTGGAGCGGGGTATGAAGAGCTTGCTTTGATTTACAGCAAGTAGATA  361
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  326  CTGAGGAACTGACGTGGAGCGGGGTATGAAGAGCTTGCTTTGATTTACAGCAAGTAGATA  385

Query  362  AATATTTGACTTGCATGAAGAGAAGCAATTTTGGGGAAGGGTTTGAATTGTTTTCTTTAA  421
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  386  AATATTTGACTTGCATGAAGAGAAGCAATTTTGGGGAAGGGTTTGAATTGTTTTCTTTAA  445

Query  422  AGATGTAATGTCCCTTTCAGAGACAGCTGATACTTCATTTaaaaaaaTCACAAAAATTTG  481
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  446  AGATGTAATGTCCCTTTCAGAGACAGCTGATACTTCATTTAAAAAAATCACAAAAATTTG  505

Query  482  AACACTGGCTAAAGATAATTGCTATTTATTTTTACAAGAAGTTTATTCTCATTTGGGAGA  541
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  506  AACACTGGCTAAAGATAATTGCTATTTATTTTTACAAGAAGTTTATTCTCATTTGGGAGA  565

Query  542  TCTGGTGATCTCCCAAGCTATCTAAAGTTTGTTANATAGCTGCATGTGGCTTTTTTAAAA  601
            |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Sbjct  566  TCTGGTGATCTCCCAAGCTATCTAAAGTTTGTTAGATAGCTGCATGTGGCTTTTTTAAAA  625

Query  602  AAGCAACAGAAACCTATCCTCACTGCCCTCCCCAGTCTCTCTTAAAGTTGGAA  654
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  626  AAGCAACAGAAACCTATCCTCACTGCCCTCCCCAGTCTCTCTTAAAGTTGGAA  678
```

Appendix 5: pMIR- *CDKN1B*-954 plasmid confirmation.

Score	Expect	Identities	Gaps	Strand	Frame
2165 bits(1172)	0.0()	1180/1184(99%)	1/1184(0%)	Plus/Plus	

Features:

```
Query  1     TCAAACGTAAACAGCTCGAATTAAGAAANATGTTTCCTTGTTTATCAGATACATCACTGCT  60
             ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Sbjct  1158  TCAAACGTAAACAGCTCGAATTAAGAATATGTTTCCTTGTTTATCAGATACATCACTGCT  1217

Query  61    TGATGAAGCAAGGAAGATATACATGAAAATTTTAAAAATACATATCGCTGACTTCATGGA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1218  TGATGAAGCAAGGAAGATATACATGAAAATTTTAAAAATACATATCGCTGACTTCATGGA  1277

Query  121   ATGGACATCCTGTATAAGCACTGAAAAANCAACACCCCAATAACACTAAAATTTTAGGCA  180
             |||||||||||||||||||||||||||| ||||| | ||||||||||||||||||||||
Sbjct  1278  ATGGACATCCTGTATAAGCACTGAAAAA-CAACAACACAATAACACTAAAATTTTAGGCA  1336

Query  181   CTCTTAAATGATCTGCCTCTAAAAGCGTTGGATGTAGCATTATGCAATTAGGTTTTTCCT  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1337  CTCTTAAATGATCTGCCTCTAAAAGCGTTGGATGTAGCATTATGCAATTAGGTTTTTCCT  1396

Query  241   TATTTGCTTCATTGTACTACCTGTGTATATAGTTTTTACCTTTTATGTAGCACATAAACT  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1397  TATTTGCTTCATTGTACTACCTGTGTATATAGTTTTTACCTTTTATGTAGCACATAAACT  1456

Query  301   TTGGGGAAGGGAGGGCAGGGTGGGGCTGAGGAACTGACGTGGAGCGGGGTATGAAGAGCT  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1457  TTGGGGAAGGGAGGGCAGGGTGGGGCTGAGGAACTGACGTGGAGCGGGGTATGAAGAGCT  1516

Query  361   TGCTTTGATTTACAGCAAGTAGATAAATATTTGACTTGCATGAAGAGAAGCAATTTTGGG  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1517  TGCTTTGATTTACAGCAAGTAGATAAATATTTGACTTGCATGAAGAGAAGCAATTTTGGG  1576

Query  421   GAAGGGTTTGAATTGTTTTCTTTAAAGATGTAATGTCCCTTTCAGAGACAGCTGATACTT  480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1577  GAAGGGTTTGAATTGTTTTCTTTAAAGATGTAATGTCCCTTTCAGAGACAGCTGATACTT  1636

Query  481   CATTTaaaaaaaTCACAAAAATTTGAACACTGGCTAAAGATAATTGCTATTTATTTTTAC  540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1637  CATTTAAAAAAATCACAAAAATTTGAACACTGGCTAAAGATAATTGCTATTTATTTTTAC  1696

Query  541   AAGAAGTTTATTCTCATTTGGGAGATCTGGTGATCTCCCAAGCTATCTAAAGTTTGTTAG  600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1697  AAGAAGTTTATTCTCATTTGGGAGATCTGGTGATCTCCCAAGCTATCTAAAGTTTGTTAG  1756

Query  601   ATAGCTGCATGTGGCTTTTTTAAAAAAGCAACAGAAACCTATCCTCACTGCCCTCCCCAG  660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1757  ATAGCTGCATGTGGCTTTTTTAAAAAAGCAACAGAAACCTATCCTCACTGCCCTCCCCAG  1816

Query  661   TCTCTCTTAAAGTTGGAATTTACCAGTTAATTACTCAGCAGAATGGTGATCACTCCAGGT  720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1817  TCTCTCTTAAAGTTGGAATTTACCAGTTAATTACTCAGCAGAATGGTGATCACTCCAGGT  1876

Query  721   AGTTTGGGGCAAAAATCCGAGGTGCTTGGGAGTTTTGAATGTTAAGAATTGACCATCTGC  780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1877  AGTTTGGGGCAAAAATCCGAGGTGCTTGGGAGTTTTGAATGTTAAGAATTGACCATCTGC  1936

Query  781   TTTTATTAAATTTGTTGACAAAATTTTCTCATTTTCTTTTCACTTCGGGCTGTGTAAACA  840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1937  TTTTATTAAATTTGTTGACAAAATTTTCTCATTTTCTTTTCACTTCGGGCTGTGTAAACA  1996

Query  841   CAGTCAAAATAATTCTAAATCCCTCGATATTTTTAAAGATCTGTAAGTAACTTCACATTA  900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1997  CAGTCAAAATAATTCTAAATCCCTCGATATTTTTAAAGATCTGTAAGTAACTTCACATTA  2056

Query  901   AAAAATGAAATATTTTTTAATTTAAAGCTTACTCTGTCCATTTATCCACAGGAAAGTGTT  960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2057  AAAAATGAAATATTTTTTAATTTAAAGCTTACTCTGTCCATTTATCCACAGGAAAGTGTT  2116

Query  961   ATTTTTCAAGGAAGGTTCATGTAGAGAAAAGCACACTTGTAGGATAAGTGAAATGGATAC  1020
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2117  ATTTTTCAAGGAAGGTTCATGTAGAGAAAAGCACACTTGTAGGATAAGTGAAATGGATAC  2176

Query  1021  TACATCTTTAAACAGTATTTCATTGCCTGTGTATGGAAAAACCATTTGAAGTGTACCTGT  1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2177  TACATCTTTAAACAGTATTTCATTGCCTGTGTATGGAAAAACCATTTGAAGTGTACCTGT  2236

Query  1081  GTACATAACTCTGTAAAAACACTGAAAAATTATACTAACTTATTTATGTTAAAAGAtttt  1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2237  GTACATAACTCTGTAAAAACACTGAAAAATTATACTAACTTATTTATGTTAAAAGATTTT  2296

Query  1141  ttttAATCTAGACAATATACAAGCCAAAGTGGCATGTTTTGTGC  1184
             |||||||||||||||||||||||||||||||||||||||||||
Sbjct  2297  TTTTAATCTAGACAATATACAAGCCAAAGTGGCATGTTTTGTGC  2340
```

Appendix 6: pMIR-*CDKN1B*-1400 plasmid confirmation.

Score	Expect	Identities	Gaps	Strand	Frame
5943 bits(3218)	0.0()	3263/3281(99%)	18/3281(0%)	Plus/Plus	

Features:

```
Query  1     AGAGCCTAAGGATAATGTCTGGCATAGAGGAGGCACTCAATAAATATTTGTTGAAAGAAA  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1691  AGAGCCTAAGGATAATGTCTGGCATAGAGGAGGCACTCAATAAATATTTGTTGAAAGAAA  1750

Query  61    GAAGGAATAAGCAAATTAATAGAAGGTGCTCCGTAAATATTTGTGAAATTGAATTAAAAG  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1751  GAAGGAATAAGCAAATTAATAGAAGGTGCTCCGTAAATATTTGTGAAATTGAATTAAAAG  1810

Query  121   TGTCAGAAGGAGATGACTGTGAAATTCCCAAAGTCTAAAAATATGTGAGCTTAGAAGAAT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1811  TGTCAGAAGGAGATGACTGTGAAATTCCCAAAGTCTAAAAATATGTGAGCTTAGAAGAAT  1870

Query  181   GGTGGAGTTGAGTGCTGCCTCA--------ATCCTCGAGGAAGGACTGAAACTGTGTGCT  232
             ||||||||||||||||||||||        ||||||||||||||||||||||||||||||
Sbjct  1871  GGTGGAGTTGAGTGCTGCCTCATTATAAACATCCTCGAGGAAGGACTGAAACTGTGTGCT  1930

Query  233   TGCGGTGGGAGGGGCAGCTGGGCAAGGAACCGTGAACCTTCGCAGAAACATTTGGGGCTG  292
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1931  TGCGGTGGGAGGGGCAGCTGGGCAAGGAACCGTGAACCTTCGCAGAAACATTTGGGGCTG  1990

Query  293   CAGAACTTGGGTGAGAGCGCTGCATCTGGGAGCTGGCGACGCTGGCGGCTTGCTCATTCA  352
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1991  CAGAACTTGGGTGAGAGCGCTGCATCTGGGAGCTGGCGACGCTGGCGGCTTGCTCATTCA  2050

Query  353   CCCCATCTGAACACTTGTCTATGACACAGGTGTTTTCTCTTAAGTTATTTTGGTCTTTGC  412
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2051  CCCCATCTGAACACTTGTCTATGACACAGGTGTTTTCTCTTAAGTTATTTTGGTCTTTGC  2110

Query  413   CTCTCTCCTCAGGTTGTGAAGATTACAGAAATCTGGGATGGCTTATGGGACGCTTCTCAG  472
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2111  CTCTCTCCTCAGGTTGTGAAGATTACAGAAATCTGGGATGGCTTATGGGACGCTTCTCAG  2170

Query  473   CCCTAAGTAGGAAAACAGCAGTGAAAATGGCAACCAAAACATCACGCAGGACTGGGGGTT  532
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2171  CCCTAAGTAGGAAAACAGCAGTGAAAATGGCAACCAAAACATCACGCAGGACTGGGGGTT  2230

Query  533   TTGGGGAAACAGCTCACTTTAGAGCAGTGCAGTGTAGAGCTTTCCGTCTTTTACCAGGGT  592
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2231  TTGGGGAAACAGCTCACTTTAGAGCAGTGCAGTGTAGAGCTTTCCGTCTTTTACCAGGGT  2290

Query  593   CCACCTTTAACACTGTTTATCTGAAAATTTTCCCCCTGGCTTACTCGCTTGCAGCTGCCC  652
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2291  CCACCTTTAACACTGTTTATCTGAAAATTTTCCCCCTGGCTTACTCGCTTGCAGCTGCCC  2350

Query  653   ACTTTGCAGAAGGATGGCGCTCTGATCTCTACGCTCCCTGTTCCTTCAGGGACTCCATAG  712
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2351  ACTTTGCAGAAGGATGGCGCTCTGATCTCTACGCTCCCTGTTCCTTCAGGGACTCCATAG  2410

Query  713   TAttttttttCACGCGTCGTCGCTACTACAGCAGACGCCTGCGTTCTCATTATTTGCTGT  772
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2411  TATTTTTTTTCACGCGTCGTCGCTACTACAGCAGACGCCTGCGTTCTCATTATTTGCTGT  2470

Query  773   ACAGATCTCCGGTGCCTTGACTGTAAACAAAACACTTTAGATCATTGTGAGGTCGATGTA  832
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2471  ACAGATCTCCGGTGCCTTGACTGTAAACAAAACACTTTAGATCATTGTGAGGTCGATGTA  2530

Query  833   AGCACAGCCTTTCTGCTGGCAGCCAGACTTCTTAAGGTGGTGTGACTGTGACTTGCTTAC  892
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2531  AGCACAGCCTTTCTGCTGGCAGCCAGACTTCTTAAGGTGGTGTGACTGTGACTTGCTTAC  2590

Query  893   TTTTCGAGATCAACAACAACAAAGCGACAAAATGGTGCTCCTACATATTAGTTGAAAGAT  952
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2591  TTTTCGAGATCAACAACAACAAAGCGACAAAATGGTGCTCCTACATATTAGTTGAAAGAT  2650

Query  953   TCAGCATGTGAAGGGGATCGAAGTGTTTATTTTCCACTTCCATATAAGACATGAATTCCA  1012
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2651  TCAGCATGTGAAGGGGATCGAAGTGTTTATTTTCCACTTCCATATAAGACATGAATTCCA  2710

Query  1013  TGAGTAAAATCAAACTTCTGTGGCAAGGTGAACTACTCTAGAATGTCTCCATTTACATAC  1072
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2711  TGAGTAAAATCAAACTTCTGTGGCAAGGTGAACTACTCTAGAATGTCTCCATTTACATAC  2770

Query  1073  ATGTGGTAGTTTGGGTGTTTATGCATATGGATAGATGCACATATATAGAGTTCCTGTGTT  1132
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2771  ATGTGGTAGTTTGGGTGTTTATGCATATGGATAGATGCACATATATAGAGTTCCTGTGTT  2830

Query  1133  GTCTAGCAATTGTTTTAAAATTTGGACAATTATCTAATTTCTAGGGTAAGGTATAAATTA  1192
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2831  GTCTAGCAATTGTTTTAAAATTTGGACAATTATCTAATTTCTAGGGTAAGGTATAAATTA  2890
```

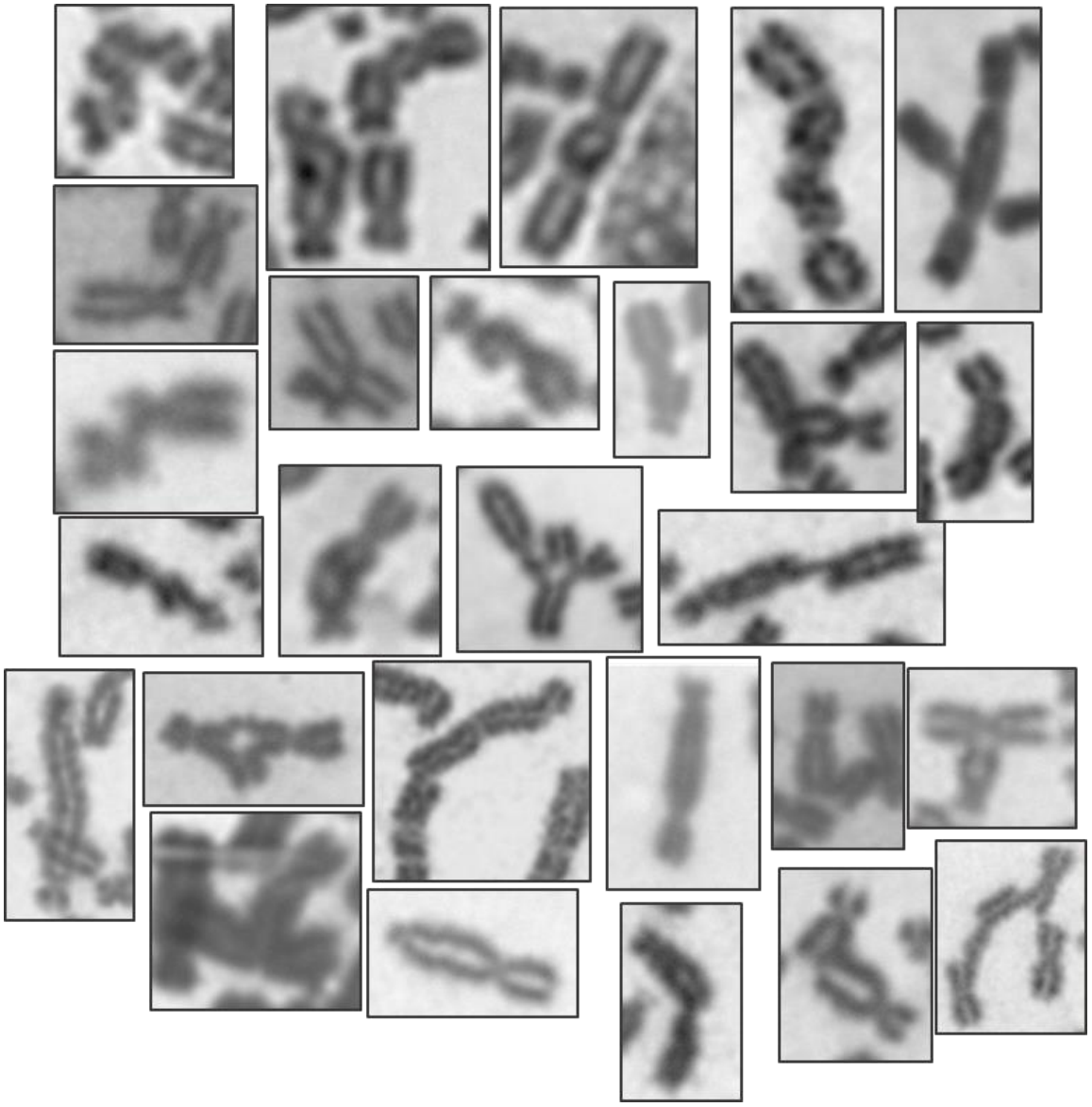
Appendix 7: pGL3-*CDKN1B* promoter-3500 plasmid confirmation.

```
Query  1193  TGGTAGGGAGGCCTACCCTAATTTTCCTGTTCCTTTTCCCCCAGTCTGCAGTCCAATAAA  1252
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2891  TGGTAGGGAGGCCTACCCTAATTTTCCTGTTCCTTTTCCCCCAGTCTGCAGTCCAATAAA  2950

Query  1253  TTGACAGCCTTAAAAGTAGAAAAACTAAAGAGGATGAGACCTCTTGCTTGATCCTAGGTG  1312
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  2951  TTGACAGCCTTAAAAGTAGAAAAACTAAAGAGGATGAGACCTCTTGCTTGATCCTAGGTG  3010

Query  1313  AATTCTTTTCTGTCAGTTAGGTAGGAAGTCCTGACTTGAAAACTAGTTCTGGGCACTGCC  1372
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3011  AATTCTTTTCTGTCAGTTAGGTAGGAAGTCCTGACTTGAAAACTAGTTCTGGGCACTGCC  3070

Query  1373  CCCTTTACTGTTCTCTGGGTATCAACCCCTGTCCTTCAATTTTAGTTGAACTAGTGGATG  1432
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3071  CCCTTTACTGTTCTCTGGGTATCAACCCCTGTCCTTCAATTTTAGTTGAACTAGTGGATG  3130

Query  1433  GTGATACCACAGGCTCAAGACAGCTGCATTTAAATATCAGTGACCACAGGCCACATCAAG  1492
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3131  GTGATACCACAGGCTCAAGACAGCTGCATTTAAATATCAGTGACCACAGGCCACATCAAG  3190

Query  1493  GAAACATCTGCAGGCAACCCAGGGCCTGGGAAGGAGCCATTTTCAGTCACTTGTAAGACA  1552
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3191  GAAACATCTGCAGGCAACCCAGGGCCTGGGAAGGAGCCATTTTCAGTCACTTGTAAGACA  3250

Query  1553  GCAGGACCTGCAGACTACAGCACAATCAAACTCAGACAAAACCCTGAACCAGTGAGAACC  1612
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3251  GCAGGACCTGCAGACTACAGCACAATCAAACTCAGACAAAACCCTGAACCAGTGAGAACC  3310

Query  1613  ATTAGGAAGGAAAGGAACAGAAAATGAACCAACCTGAGTGTTAGGAGACTTGCATCTAGT  1672
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3311  ATTAGGAAGGAAAGGAACAGAAAATGAACCAACCTGAGTGTTAGGAGACTTGCATCTAGT  3370

Query  1673  CCTGACTCCGGTACCAACCGAATGCATGTCCCTGGACAGGAAACCTCTCTGAGTCTCGAT  1732
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3371  CCTGACTCCGGTACCAACCGAATGCATGTCCCTGGACAGGAAACCTCTCTGAGTCTCGAT  3430

Query  1733  TTCCTCCGTGGTAAAAAGGAGAGGGTTAAACCACAGGGTCCCGAGGGTCCCTTCCAGCTG  1792
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3431  TTCCTCCGTGGTAAAAAGGAGAGGGTTAAACCACAGGGTCCCGAGGGTCCCTTCCAGCTG  3490

Query  1793  TCACATTCTGGAGCGTATGAGATGAGGTAGGCACACAAAGTGGACAAGATGTGGCTAAGA  1852
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3491  TCACATTCTGGAGCGTATGAGATGAGGTAGGCACACAAAGTGGACAAGATGTGGCTAAGA  3550

Query  1853  AAACAAGCTACACATCAAGCTCATCTGTAGCATAGGTGCTTAAGAAAACTTTGCTGCTGT  1912
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3551  AAACAAGCTACACATCAAGCTCATCTGTAGCATAGGTGCTTAAGAAAACTTTGCTGCTGT  3610

Query  1913  GTAATATTAGAACGGAAGGTTGGTTTCCAGTAAAATGCATTAACTTTGGCTCAAACCAAG  1972
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3611  GTAATATTAGAACGGAAGGTTGGTTTCCAGTAAAATGCATTAACTTTGGCTCAAACCAAG  3670

Query  1973  ATGATGGGTACCGGGCATGGGGGTGGGGAGGCAGTTGAAGATCCACTGAGCTTTGTCTCA  2032
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3671  ATGATGGGTACCGGGCATGGGGGTGGGGAGGCAGTTGAAGATCCACTGAGCTTTGTCTCA  3730

Query  2033  GGGCAGCCCTGCTCATCGTCCTACTTTACCTTCCACCACGGTGCT-----CCACACTGAG  2087
             ||||||||||||||||||||||||||||||||||||||||||||||     ||||||||||
Sbjct  3731  GGGCAGCCCTGCTCATCGTCCTACTTTACCTTCCACCACGGTGCTCAAGCCCACACTGAG  3790

Query  2088  AGAGAAATTTCCAGCTGCAAAAGGGAGAAGAGAAACGCTGGAATACTAGTATCGGACGTT  2147
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3791  AGAGAAATTTCCAGCTGCAAAAGGGAGAAGAGAAACGCTGGAATACTAGTATCGGACGTT  3850

Query  2148  AGGACATGGTTGTGGTGTTTTAAAAATCATTTCATCATCTGGAGTTTGACCCCGAGGGGA  2207
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3851  AGGACATGGTTGTGGTGTTTTAAAAATCATTTCATCATCTGGAGTTTGACCCCGAGGGGA  3910

Query  2208  GTATTTTCACCCTTCAGCCCTCTGAAAGCATTCACTAGCATCTGAATATTGTTCTGAGTT  2267
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3911  GTATTTTCACCCTTCAGCCCTCTGAAAGCATTCACTAGCATCTGAATATTGTTCTGAGTT  3970

Query  2268  TGTTGGAGCAGTGAAATCTGGTGAGAGAGAAGGGTGGAGGAAGGAAGGAGCTGTTGTATT  2327
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  3971  TGTTGGAGCAGTGAAATCTGGTGAGAGAGAAGGGTGGAGGAAGGAAGGAGCTGTTGTATT  4030

Query  2328  TGGCGGCTGGACTCAGGTAGAGGAAACTGCTACAATCCCGGGAAAGAACAGAAAAGTAGA  2387
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4031  TGGCGGCTGGACTCAGGTAGAGGAAACTGCTACAATCCCGGGAAAGAACAGAAAAGTAGA  4090

Query  2388  AAGGGACGAGTTCCCACACGCAGCCAATGTCCATGGCCTTAACTGTGCTTGGGAAGGAAG  2447
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4091  AAGGGACGAGTTCCCACACGCAGCCAATGTCCATGGCCTTAACTGTGCTTGGGAAGGAAG  4150

Query  2448  ATCCTGGGCCAGGGGTGTACCCTCGTTTTTCAAAACTAAACGTGTCTGAGACAGCTACAA  2507
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4151  ATCCTGGGCCAGGGGTGTACCCTCGTTTTTCAAAACTAAACGTGTCTGAGACAGCTACAA  4210
```

Appendix 7: pGL3-*CDKN1B* promoter-3500 plasmid confirmation.

```
Query  2508  AGTTTATTAAGGGACTTGAGAGACTAGAGttttttgtttttttttttttAATCTTGAGTTC  2567
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4211  AGTTTATTAAGGGACTTGAGAGACTAGAGTTTTTTGTTTTTTTTTTTTTAATCTTGAGTTC  4270

Query  2568  CTTTCTTATTTTCATTGAGGGAGAGCTTGAGTTCATGATAAGTGCCGCGTCTACTCCTGG  2627
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4271  CTTTCTTATTTTCATTGAGGGAGAGCTTGAGTTCATGATAAGTGCCGCGTCTACTCCTGG  4330

Query  2628  CTAATTTCTAAAAGAAAGACGTTCGCTTTGGCTTCTTCCCTAGGCCCCCAGCCTCCCCAG  2687
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4331  CTAATTTCTAAAAGAAAGACGTTCGCTTTGGCTTCTTCCCTAGGCCCCCAGCCTCCCCAG  4390

Query  2688  GGATGGCAGAAACTTCTGGGTTAAGGCTGAGCGAACCATTGCCCACTGCCTCCACCAGCC  2747
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4391  GGATGGCAGAAACTTCTGGGTTAAGGCTGAGCGAACCATTGCCCACTGCCTCCACCAGCC  4450

Query  2748  CCCAGCAAAGGCACGCCGGCgggggggCGCCCAGccccccccAGCAAACGCTCCGCGGCCT  2807
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4451  CCCAGCAAAGGCACGCCGGCGGGGGGGCGCCCAGCCCCCCCAGCAAACGCTCCGCGGCCT  4510

Query  2808  CCCCCGCAGACCACGAGGTGGGGGCCGCTGGGGAGGGCCGAGCTGGGGGCAGCTCGCCAC  2867
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4511  CCCCCGCAGACCACGAGGTGGGGGCCGCTGGGGAGGGCCGAGCTGGGGGCAGCTCGCCAC  4570

Query  2868  CCCGGCTCCTAGCGAGCTGCCGGCGACCTTCGCGGTCCTCTGGTCCAGGTCCCGGCTTCC  2927
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4571  CCCGGCTCCTAGCGAGCTGCCGGCGACCTTCGCGGTCCTCTGGTCCAGGTCCCGGCTTCC  4630

Query  2928  CGGGCGAGGAGCGGGAGGGAGGTCGGGGCTTAGGCGCCGCGGCGAACCCGCCAACGCAGC  2987
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4631  CGGGCGAGGAGCGGGAGGGAGGTCGGGGCTTAGGCGCCGCGGCGAACCCGCCAACGCAGC  4690

Query  2988  GCCGGGCCCCGAACCTCAGGCCCCGCCCCAGGTTCCCGGCCGTTTGGCTAGTTTGTTTGT  3047
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4691  GCCGGGCCCCGAACCTCAGGCCCCGCCCCAGGTTCCCGGCCGTTTGGCTAGTTTGTTTGT  4750

Query  3048  CTTAATTTTAATTTCTCCGAGGCCAGCCAGAGCAGGTTTGTTGGCAGCAGTACCCCTCCA  3107
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4751  CTTAATTTTAATTTCTCCGAGGCCAGCCAGAGCAGGTTTGTTGGCAGCAGTACCCCTCCA  4810

Query  3108  GCAGTCACGCGACCAGCCAATCTCCCGGCGGCGCTC-----GGCGGCGCGCTCGGGAACG  3162
             ||||||||||||||||||||||||||||||||||||     ||||||||||||||||||||
Sbjct  4811  GCAGTCACGCGACCAGCCAATCTCCCGGCGGCGCTCGGGGAGGCGGCGCGCTCGGGAACG  4870

Query  3163  AGGGGAGGTGGCGGAACCGCGCCGGGGCCACCTTAAGGCCGCGCTCGCCAGCCTCGGCGG  3222
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  4871  AGGGGAGGTGGCGGAACCGCGCCGGGGCCACCTTAAGGCCGCGCTCGCCAGCCTCGGCGG  4930

Query  3223  GGCGGCTCCCGCCGCCGCAACCAATGGATCTCCTCCTCTGT  3263
             |||||||||||||||||||||||||||||||||||||||||
Sbjct  4931  GGCGGCTCCCGCCGCCGCAACCAATGGATCTCCTCCTCTGT  4971
```

Appendix 7: pGL3-*CDKN1B* promoter-3500 plasmid confirmation.



Appendix 8: examples of chromatid/chromatin fusions observed in the various karyotypes analysed.

References

Aaltonen, L., Peltomaki, P., Leach, F., Sistonen, P., Pylkkanen, L., Mecklin, J., Jarvinen, H., Powell, S., Jen, J., Hamilton*, et al.* (1993). Clues to the pathogenesis of familial colorectal cancer. Science *260*, 812-816.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA *100*.

Alam, J., and Cook, J.L. (1990). Reporter genes: application to the study of mammalian gene transcription. Anal Biochem *188*, 245-254.

Ali, R., Rakha, E.A., Madhusudan, S., and Bryant, H.E. (2017). DNA damage repair in breast cancer and its therapeutic implications. Pathology *49*, 156-165.

Alkarain, A., and Slingerland, J. (2004). Deregulation of p27 by oncogenic signaling and its prognostic significance in breast cancer. Breast Cancer Research *6*, 13-21.

Ambudkar, S.V., Kimchi-Sarfaty, C., Sauna, Z.E., and Gottesman, M.M. (2002). P-glycoprotein: from genomics to mechanism. Oncogene *22*, 7468.

Andersen, H., Mejlvang, J., Mahmood, S., Gromova, I., Gromov, P., Lukanidin, E., Kriajevska, M., Mellon, J.K., and Tulchinsky, E. (2005). Immediate and delayed effects of E-cadherin inhibition on gene regulation and cell motility in human epidermoid carcinoma cells. Mol Cell Biol *25*, 9138-9150.

Arakawa, H., Bednar, T., Wang, M., Paul, K., Mladenov, E., Bencsik-Theilen, A.A., and Iliakis, G. (2012). Functional redundancy between DNA ligases I and III in DNA replication in vertebrate cells. Nucleic Acids Res *40*, 2599-2610.

Arakawa, H., and Iliakis, G. (2015). Alternative Okazaki Fragment Ligation Pathway by DNA Ligase III. Genes (Basel) *6*, 385-398.

Artandi, S.E., and DePinho, R.A. (2010). Telomeres and telomerase in cancer. Carcinogenesis *31*, 9-18.

Bachs, O., Gallastegui, E., Orlando, S., Bigas, A., Morante-Redolat, J.M., Serratosa, J., Farinas, I., Aligue, R., and Pujol, M.J. (2018). Role of p27(Kip1) as a transcriptional regulator. Oncotarget *9*, 26259-26278.

Bakhoum, S.F., Ngo, B., Laughney, A.M., Cavallo, J.-A., Murphy, C.J., Ly, P., Shah, P., Sriram, R.K., Watkins, T.B.K., Taunk, N.K.*, et al.* (2018). Chromosomal instability drives metastasis through a cytosolic DNA response. Nature *553*, 467.

Ballas, N., Battaglioli, E., Atouf, F., Andres, M.E., Chenoweth, J., Anderson, M.E., Burger, C., Moniwa, M., Davie, J.R., Bowers, W.J.*, et al.* (2001). Regulation of Neuronal Traits by a Novel Transcriptional Complex. Neuron *31*, 353-365.

Barbash, O., and Alan Diehl, J. (2008). Chapter 13 - Regulation of the Cell Cycle. In The Molecular Basis of Cancer (Third Edition), J. Mendelsohn, P.M. Howley, M.A. Israel, J.W. Gray, and C.B. Thompson, eds. (Philadelphia: W.B. Saunders), pp. 177-188.

Bassez, G., Camand, O.J., Cacheux, V., Kobetz, A., Dastot-Le Moal, F., Marchant, D., Catala, M., Abitbol, M., and Goossens, M. (2004). Pleiotropic and diverse expression of ZFHX1B gene transcripts during mouse and human development supports the various clinical manifestations of the "Mowat-Wilson" syndrome. Neurobiol Dis *15*, 240-250.

Basu, A. (2018). DNA Damage, Mutagenesis and Cancer. International Journal of Molecular Sciences *19*, 970.

Belletti, B., and Baldassarre, G. (2012). New light on p27(kip1) in breast cancer. Cell Cycle *11*, 3701-3702.

Bencivenga, D., Caldarelli, I., Stampone, E., Mancini, F.P., Balestrieri, M.L., Della Ragione, F., and Borriello, A. (2017). p27Kip1 and human cancers: A reappraisal of a still enigmatic protein. Cancer Letters *403*, 354-365.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K.*, et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315-326.

Berton, S., Cusan, M., Segatto, I., Citron, F., D'Andrea, S., Benevol, S., Avanzo, M., Dall'Acqua, A., Schiappacassi, M., Bristow, R.G.*, et al.* (2017). Loss of p27kip1 increases genomic instability and induces radio-resistance in luminal breast cancer cells. Scientific Reports *7*, 595.

Berx, G., Cleton-Jansen, A.M., Strumane, K., de Leeuw, W.J., Nollet, F., van Roy, F., and Cornelisse, C. (1996). E-cadherin is inactivated in a majority of invasive human lobular breast cancers by truncation mutations throughout its extracellular domain. Oncogene *13*, 1919-1925.

Besson, A., Assoian, R.K., and Roberts, J.M. (2004a). Regulation of the cytoskeleton: an oncogenic function for CDK inhibitors? Nat Rev Cancer *4*, 948-955.

Besson, A., Gurian-West, M., Schmidt, A., Hall, A., and Roberts, J.M. (2004b). p27Kip1 modulates cell migration through the regulation of RhoA activation. Genes Dev *18*, 862-876.

Besson, A., Hwang, H.C., Cicero, S., Donovan, S.L., Gurian-West, M., Johnson, D., Clurman, B.E., Dyer, M.A., and Roberts, J.M. (2007). Discovery of an oncogenic activity in p27Kip1 that causes stem cell expansion and a multiple tumor phenotype. Genes Dev *21*, 1731-1746.

Bhowmick, N.A., Neilson, E.G., and Moses, H.L. (2004). Stromal fibroblasts in cancer initiation and progression. Nature *432*.

Blanco, M.J., Moreno-Bueno, G., Sarrio, D., Locascio, A., Cano, A., Palacios, J., and Nieto, M.A. (2002). Correlation of Snail expression with histological grade and lymph node status in breast carcinomas. Oncogene *21*, 3241.

Blasco, M.A. (2005). Telomeres and human disease: ageing, cancer and beyond. Nat Rev Genet *6*, 611-622.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nature Medicine *3*, 730.

Brabletz, T. (2012). EMT and MET in Metastasis: Where Are the Cancer Stem Cells? Cancer Cell *22*, 699-701.

Brabletz, T., Kalluri, R., Nieto, M.A., and Weinberg, R.A. (2018). EMT in cancer. Nature Reviews Cancer *18*, 128.

Cai, Z., Cao, Y., Luo, Y., Hu, H., and Ling, H. (2018). Signalling mechanism(s) of epithelial–mesenchymal transition and cancer stem cells in tumour therapeutic resistance. Clinica Chimica Acta *483*, 156-163.

Campeau, P.M., Foulkes, W.D., and Tischkowitz, M.D. (2008). Hereditary breast cancer: new genetic developments, new therapeutic avenues. Human Genetics *124*, 31-42.

Carver, E.A., Jiang, R., Lan, Y., Oram, K.F., and Gridley, T. (2001). The Mouse Snail Gene Encodes a Key Regulator of the Epithelial-Mesenchymal Transition. Molecular and Cellular Biology *21*, 8184-8188.

Casimiro, M.C., Crosariol, M., Loro, E., Ertel, A., Yu, Z., Dampier, W., Saria, E.A., Papanikolaou, A., Stanek, T.J., Li, Z.*, et al.* (2012). ChIP sequencing of cyclin D1 reveals a transcriptional role in chromosomal instability in mice. J Clin Invest *122*, 833-843.

Casimiro, M.C., Velasco-Velazquez, M., Aguirre-Alvarado, C., and Pestell, R.G. (2014). Overview of cyclins D1 function in cancer and the CDK inhibitor landscape: past and present. Expert Opin Investig Drugs *23*, 295-304.

Castells, M., Thibault, B., Delord, J.P., and Couderc, B. (2012). Implication of tumor microenvironment in chemoresistance: tumor-associated stromal cells protect tumor cells from cell death. Int J Mol Sci *13*, 9545-9571.

Cavallaro, U., and Christofori, G. (2004). Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat Rev Cancer *4*, 118-132.

Ceccaldi, R., Sarangi, P., and D'Andrea, A.D. (2016). The Fanconi anaemia pathway: new players and new functions. Nature Reviews Molecular Cell Biology *17*, 337.

Chang, B.-l., Zheng, S.L., Isaacs, S.D., Wiley, K.E., Turner, A., Li, G., Walsh, P.C., Meyers, D.A., Isaacs, W.B., and Xu, J. (2004). A Polymorphism in the CDKN1B Gene Is Associated with Increased Risk of Hereditary Prostate Cancer. Cancer Research *64*, 1997-1999.

Chen, H., Mo, X., Yu, J., Huang, S., Huang, Z., and Gao, L. (2014). Interference of Skp2 effectively inhibits the development and metastasis of colon carcinoma. Mol Med Rep *10*, 1129-1135.

Chiarle, R., Budel, L.M., Skolnik, J., Frizzera, G., Chilosi, M., Corato, A., Pizzolo, G., Magidson, J., Montagnoli, A., Pagano, M.*, et al.* (2000). Increased proteasome degradation of cyclin-dependent kinase inhibitor p27 is associated with a decreased overall survival in mantle cell lymphoma. Blood *95*, 619-626.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem *162*, 156-159.

Chrzanowska, K.H., Gregorek, H., Dembowska-Bagińska, B., Kalina, M.A., and Digweed, M. (2012). Nijmegen breakage syndrome (NBS). Orphanet Journal of Rare Diseases *7*, 13-13.

Chu, I.M., Hengst, L., and Slingerland, J.M. (2008). The Cdk inhibitor p27 in human cancer: prognostic potential and relevance to anticancer therapy. Nature Reviews Cancer *8*, 253.

Cobrinik, D. (2005). Pocket proteins and cell cycle control. Oncogene *24*, 2796.

Cojoc, M., Mäbert, K., Muders, M.H., and Dubrovska, A. (2015). A role for cancer stem cells in therapy resistance: Cellular and molecular mechanisms. Seminars in Cancer Biology *31*, 16-27.

Coller, H.A. (2007). What's taking so long? S-phase entry from quiescence versus proliferation. Nature Reviews Molecular Cell Biology *8*, 667.

Comaills, V., Kabeche, L., Morris, R., Buisson, R., Yu, M., Madden, M.W., LiCausi, J.A., Boukhali, M., Tajima, K., Pan, S.*, et al.* (2016). Genomic Instability Is Induced by Persistent Proliferation of Cells Undergoing Epithelial-to-Mesenchymal Transition. Cell Reports *17*, 2632-2647.

D'Andrea, A.D. (2015). 4 - DNA Repair Pathways and Human Cancer. In The Molecular Basis of Cancer (Fourth Edition), J. Mendelsohn, J.W. Gray, P.M. Howley, M.A. Israel, and C.B. Thompson, eds. (Philadelphia: Content Repository Only!), pp. 47-66.e42.

Darling, D., P. Stearman, R., Qi, Y., Qiu, M.-S., and P. Feller, J. (2003). Expression of Zfhep/δEF1 protein in palate, neural progenitors, and differentiated neurons, Vol 3.

Dave, N., Guaita-Esteruelas, S., Gutarra, S., Frias, À., Beltran, M., Peiró, S., and de Herreros, A.G. (2011). Functional Cooperation between Snail1 and Twist in the Regulation of ZEB1 Expression during Epithelial to Mesenchymal Transition. The Journal of Biological Chemistry *286*, 12024-12032.

Davis, A.J., and Chen, D.J. (2013). DNA double strand break repair via non-homologous end-joining. Translational cancer research *2*, 130-143.

De Craene, B., and Berx, G. (2013). Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer *13*, 97-110.

De Craene, B., Denecker, G., Vermassen, P., Taminau, J., Mauch, C., Derore, A., Jonkers, J., Fuchs, E., and Berx, G. (2013). Epidermal Snail expression drives skin cancer initiation and progression through enhanced cytoprotection, epidermal stem/progenitor cell expansion and enhanced metastatic potential. Cell Death And Differentiation *21*, 310.

De Smedt, L., Palmans, S., Andel, D., Govaere, O., Boeckx, B., Smeets, D., Galle, E., Wouters, J., Barras, D., Suffiotti, M.*, et al.* (2017). Expression profiling of budding cells in colorectal cancer reveals an EMT-like phenotype and molecular subtype switching. British journal of cancer *116*, 58-65.

Deng, C.-X. (2006). BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. Nucleic Acids Research *34*, 1416-1426.

Dong, C., Wu, Y., Wang, Y., Wang, C., Kang, T., Rychahou, P.G., Chi, Y.I., Evers, B.M., and Zhou, B.P. (2012a). Interaction with Suv39H1 is critical for Snail-mediated E-cadherin repression in breast cancer. Oncogene *32*, 1351.

Dong, C., Wu, Y., Yao, J., Wang, Y., Yu, Y., Rychahou, P.G., Evers, B.M., and Zhou, B.P. (2012b). G9a interacts with Snail and is critical for Snail-mediated E-cadherin repression in human breast cancer. The Journal of Clinical Investigation *122*, 1469-1486.

Dueva, R., and Iliakis, G. (2013). Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. Translational Cancer Research *2*, 163-177.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M.D., Zimonjic, D.B., Donaher, J.L., Popescu, N.C., Hahn, W.C., and Weinberg, R.A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes Dev *15*, 50-65.

Ellenberger, T., and Tomkinson, A.E. (2008). Eukaryotic DNA Ligases: Structural and Functional Insights. Annual Review of Biochemistry *77*, 313-338.

Fan, Q.-M., Jing, Y.-Y., Yu, G.-F., Kou, X.-R., Ye, F., Gao, L., Li, R., Zhao, Q.-D., Yang, Y., Lu, Z.-H*., et al.* (2014). Tumor-associated macrophages promote cancer stem cell-like properties via transforming growth factor-beta1-induced epithelial–mesenchymal transition in hepatocellular carcinoma. Cancer Letters *352*, 160-168.

Fletcher, J.I., Haber, M., Henderson, M.J., and Norris, M.D. (2010). ABC transporters in cancer: more than just drug efflux pumps. Nature Reviews Cancer *10*, 147.

Friedberg, E.C. (2001). How nucleotide excision repair protects against cancer. Nature Reviews Cancer *1*, 22.

Friedberg, E.C. (2003a). DNA damage and repair. Nature *421*, 436.

Friedberg, E.C. (2003b). DNA damage and repair. Nature *421*, 436-440.

Funahashi, J., Sekido, R., Murai, K., Kamachi, Y., and Kondoh, H. (1993). Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis. Development *119*, 433-446.

Garrett-Engele, C.M., Tasch, M.A., Hwang, H.C., Fero, M.L., Perlmutter, R.M., Clurman, B.E., and Roberts, J.M. (2007). A Mechanism Misregulating p27 in Tumors Discovered in a Functional Genomic Screen. PLOS Genetics *3*, e219.

Géraud, C., Koch, P.S., Damm, F., Schledzewski, K., and Goerdt, S. (2014). The metastatic cycle: metastatic niches and cancer cell dissemination. JDDG: Journal der Deutschen Dermatologischen Gesellschaft *12*, 1012-1019.

Goldstein, L.J., Galski, H., Fojo, A., Willingham, M., Lai, S.-L., Gazdar, A., Pirker, R., Green, A., Crist, W., Brodeur, G.M*., et al.* (1989). Expression of Multidrug Resistance Gene in Human Cancers. JNCI: Journal of the National Cancer Institute *81*, 116-124.

Goossens, S., Vandamme, N., Van Vlierberghe, P., and Berx, G. (2017). EMT transcription factors in cancer development re-evaluated: Beyond EMT and MET. Biochimica et Biophysica Acta (BBA) - Reviews on Cancer *1868*, 584-591.

Gottesman, M.M. (2002). Mechanisms of Cancer Drug Resistance. Annual Review of Medicine *53*, 615-627.

Gstaiger, M., Jordan, R., Lim, M., Catzavelos, C., Mestan, J., Slingerland, J., and Krek, W. (2001). Skp2 is oncogenic and overexpressed in human cancers. Proceedings of the National Academy of Sciences *98*, 5043-5048.

Guo, W., Keckesova, Z., Donaher, Joana L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zürrer-Härdi, U., Bell, G.*, et al.* (2012). Slug and Sox9 Cooperatively Determine the Mammary Stem Cell State. Cell *148*, 1015-1028.

Gupta, P.B., Kuperwasser, C., Brunet, J.-P., Ramaswamy, S., Kuo, W.-L., Gray, J.W., Naber, S.P., and Weinberg, R.A. (2005). The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. Nature Genetics *37*, 1047.

Hakem, R. (2008). DNA-damage repair; the good, the bad, and the ugly. The EMBO Journal *27*, 589-605.

Han, L., Masani, S., Hsieh, C.L., and Yu, K. (2014). DNA ligase I is not essential for mammalian cell viability. Cell Rep *7*, 316-320.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Harrison, C., Ketchen, A.-M., Redhead, N.J., O'Sullivan, M.J., and Melton, D.W. (2002). Replication Failure, Genome Instability, and Increased Cancer Susceptibility in Mice with a Point Mutation in the DNA Ligase I Gene. Cancer Research *62*, 4065-4074.

Hattori, T., Isobe, T., Abe, K., Kikuchi, H., Kitagawa, K., Oda, T., Uchida, C., and Kitagawa, M. (2007). Pirh2 promotes ubiquitin-dependent degradation of the cyclin-dependent kinase inhibitor p27Kip1. Cancer Res *67*, 10789-10795.

Hay, E.D. (1995). An overview of epithelio-mesenchymal transformation. Acta Anat (Basel) *154*, 8-20.

He, W., Wang, X., Chen, L., and Guan, X. (2012). A Crosstalk Imbalance Between p27(Kip1) and Its Interacting Molecules Enhances Breast Carcinogenesis. Cancer Biotherapy & Radiopharmaceuticals *27*, 399-402.

Herranz, N., Pasini, D., Díaz, V.M., Francí, C., Gutierrez, A., Dave, N., Escrivà, M., Hernandez-Muñoz, I., Di Croce, L., Helin, K.*, et al.* (2008). Polycomb Complex 2 Is Required for E-cadherin Repression by the Snail1 Transcription Factor. Molecular and Cellular Biology *28*, 4772-4781.

Hershko, D., Bornstein, G., Ben-Izhak, O., Carrano, A., Pagano, M., Krausz, M.M., and Hershko, A. (2001). Inverse relation between levels of p27(Kip1) and of its ubiquitin ligase subunit Skp2 in colorectal carcinomas. Cancer *91*, 1745-1751.

Hill, L., Browne, G., and Tulchinsky, E. (2013). ZEB/miR-200 feedback loop: At the crossroads of signal transduction in cancer. International Journal of Cancer *132*, 745-754.

Hilmarsdottir, B., Briem, E., Bergthorsson, J., Magnusson, M., and Gudjonsson, T. (2014). Functional Role of the microRNA-200 Family in Breast Morphogenesis and Neoplasia. Genes *5*, 804.

Hoeijmakers, J.H.J. (2001). Genome maintenance mechanisms for preventing cancer. Nature *411*, 366.

Holohan, C., Van Schaeybroeck, S., Longley, D.B., and Johnston, P.G. (2013). Cancer drug resistance: an evolving paradigm. Nature Reviews Cancer *13*, 714.

Howes, T.R., and Tomkinson, A.E. (2012). DNA ligase I, the replicative DNA ligase. Subcell Biochem *62*, 327-341.

Huang, R.Y.-J., Guilford, P., and Thiery, J.P. (2012). Early events in cell adhesion and polarity during epithelial-mesenchymal transition. Journal of Cell Science *125*, 4417-4422.

Hugo, H., Ackland, M.L., Blick, T., Lawrence, M.G., Clements, J.A., Williams, E.D., and Thompson, E.W. (2007). Epithelial--mesenchymal and mesenchymal--epithelial transitions in carcinoma progression. J Cell Physiol *213*, 374-383.

Iwano, M., Plieth, D., Danoff, T.M., Xue, C., Okada, H., and Neilson, E.G. (2002). Evidence that fibroblasts derive from epithelium during tissue fibrosis. J Clin Invest *110*, 341-350.

Jackson, S.P., and Bartek, J. (2009). The DNA-damage response in human biology and disease. Nature *461*, 1071-1078.

Jamora, C., Lee, P., Kocieniewski, P., Azhar, M., Hosokawa, R., Chai, Y., and Fuchs, E. (2005). A Signaling Pathway Involving TGF-β2 and Snail in Hair Follicle Morphogenesis. PLOS Biology *3*, e11.

Janssen, A., van der Burg, M., Szuhai, K., Kops, G.J.P.L., and Medema, R.H. (2011). Chromosome Segregation Errors as a Cause of DNA Damage and Structural Chromosome Aberrations. Science *333*, 1895-1898.

Jasin, M., and Rothstein, R. (2013). Repair of strand breaks by homologous recombination. Cold Spring Harb Perspect Biol *5*, a012740.

Jorda, M., Olmeda, D., Vinyals, A., Valero, E., Cubillo, E., Llorens, A., Cano, A., and Fabra, A. (2005). Upregulation of MMP-9 in MDCK epithelial cell line in response to expression of the Snail transcription factor. J Cell Sci *118*, 3371-3385.

Kahlert, U.D., Joseph, J.V., and Kruyt, F.A.E. (2017). EMT- and MET-related processes in nonepithelial tumors: importance for disease progression, prognosis, and therapeutic opportunities. Molecular oncology *11*, 860-877.

Kalluri, R. (2009). EMT: When epithelial cells decide to become mesenchymal-like cells. The Journal of Clinical Investigation *119*, 1417-1419.

Kalluri, R., and Weinberg, R.A. (2009). The basics of epithelial-mesenchymal transition. J Clin Invest *119*, 1420-1428.

Krafchak, C.M., Pawar, H., Moroi, S.E., Sugar, A., Lichter, P.R., Mackey, D.A., Mian, S., Nairus, T., Elner, V., Schteingart, M.T.*, et al.* (2005). Mutations in TCF8 Cause

Posterior Polymorphous Corneal Dystrophy and Ectopic Expression of COL4A3 by Corneal Endothelial Cells. The American Journal of Human Genetics *77*, 694-708.

Kudo, Y., Kitajima, S., Sato, S., Miyauchi, M., Ogawa, I., and Takata, T. (2001). High expression of S-phase kinase-interacting protein 2, human F-box protein, correlates with poor prognosis in oral squamous cell carcinomas. Cancer Res *61*, 7044-7047.

KUO, L.J., and YANG, L.-X. (2008). γ-H2AX - A Novel Biomarker for DNA Double-strand Breaks. In Vivo *22*, 305-309.

Kwok, W.K., Ling, M.T., Lee, T.W., Lau, T.C., Zhou, C., Zhang, X., Chua, C.W., Chan, K.W., Chan, F.L., Glackin, C.*, et al.* (2005). Up-regulation of TWIST in prostate cancer and its implication as a therapeutic target. Cancer Res *65*, 5153-5162.

Lacy, E.R., Wang, Y., Post, J., Nourse, A., Webb, W., Mapelli, M., Musacchio, A., Siuzdak, G., and Kriwacki, R.W. (2005). Molecular Basis for the Specificity of p27 Toward Cyclin-dependent Kinases that Regulate Cell Division. Journal of Molecular Biology *349*, 764-773.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. Nature Reviews Molecular Cell Biology *15*, 178.

Le Chalony, C., Hoffschir, F., Gauthier, L.R., Gross, J., Biard, D.S., Boussin, F.D., and Pennaneach, V. (2012). Partial complementation of a DNA ligase I deficiency by DNA ligase III and its impact on cell survival and telomere stability in mammalian cells. Cell Mol Life Sci *69*, 2933-2949.

Lee, J.-K., Choi, Y.-L., Kwon, M., and Park, P.J. (2016). Mechanisms and Consequences of Cancer Genome Instability: Lessons from Genome Sequencing Studies. Annual Review of Pathology: Mechanisms of Disease *11*, 283-312.

Lei, L., and Lee, Z. (2005). Sensing, signaling, and responding to DNA damage: Organization of the checkpoint pathways in mammalian cells. Journal of Cellular Biochemistry *94*, 298-306.

Li, H., Mar, B.G., Zhang, H., Puram, R.V., Vazquez, F., Weir, B.A., Hahn, W.C., Ebert, B., and Pellman, D. (2017a). The EMT regulator ZEB2 is a novel dependency of human and murine acute myeloid leukemia. *129*, 497-508.

Li, Z., Yin, S., Zhang, L., Liu, W., and Chen, B. (2017b). Prognostic value of reduced E-cadherin expression in breast cancer: a meta-analysis. Oncotarget *8*, 16445-16455.

Liang, L., Deng, L., Nguyen, S.C., Zhao, X., Maulion, C.D., Shao, C., and Tischfield, J.A. (2008). Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. Nucleic Acids Research *36*, 3297-3310.

Lin, T., Ponn, A., Hu, X., Law, B.K., and Lu, J. (2010). Requirement of the histone demethylase LSD1 in Snai1-mediated transcriptional repression during epithelial-mesenchymal transition. Oncogene *29*, 4896.

Liu, S., Ginestier, C., Charafe-Jauffret, E., Foco, H., Kleer, C.G., Merajver, S.D., Dontu, G., and Wicha, M.S. (2008a). BRCA1 regulates human mammary stem/progenitor cell fate. Proceedings of the National Academy of Sciences *105*, 1680-1685.

Liu, Y., Peng, X., Tan, J., Darling, D.S., Kaplan, H.J., and Dean, D.C. (2008b). Zeb1 Mutant Mice as a Model of Posterior Corneal Dystrophy. Investigative ophthalmology & visual science *49*, 1843-1849.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods *25*, 402-408.

Lomonosov, M., Anand, S., Sangrithi, M., Davies, R., and Venkitaraman, A.R. (2003). Stabilization of stalled DNA replication forks by the BRCA2 breast cancer susceptibility protein. Genes & Development *17*, 3017-3022.

Long, J., Zuo, D., and Park, M. (2005). Pc2-mediated Sumoylation of Smad-interacting Protein 1 Attenuates Transcriptional Repression of E-cadherin. Journal of Biological Chemistry *280*, 35477-35489.

Lovisa, S., LeBleu, V.S., Tampe, B., Sugimoto, H., Vadnagara, K., Carstens, J.L., Wu, C.C., Hagos, Y., Burckhardt, B.C., Pentcheva-Hoang, T.*, et al.* (2015). Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. Nat Med *21*, 998-1009.

Lv, Y., Niu, Y., Li, C., Zheng, X., Geng, Q., and Han, Y. (2017). Aberrant Level of Skp2 and p27KIP1 in Intraductal Proliferative Lesions is Associated with Tumorigenesis. Cancer Investigation *35*, 414-422.

Lyons, J.G., Patel, V., Roue, N.C., Fok, S.Y., Soon, L.L., Halliday, G.M., and Gutkind, J.S. (2008). Snail Up-regulates Proinflammatory Mediators and Inhibits Differentiation in Oral Keratinocytes. *68*, 4525-4530.

Mahmood, T., and Yang, P.-C. (2012). Western Blot: Technique, Theory, and Trouble Shooting. North American Journal of Medical Sciences *4*, 429-434.

Malumbres, M., and Barbacid, M. (2001). To cycle or not to cycle: a critical decision in cancer. Nature Reviews Cancer *1*, 222.

Malumbres, M., and Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. Nat Rev Cancer *9*, 153-166.

Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M.*, et al.* (2008). The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. Cell *133*, 704-715.

Marcucci, F., Stassi, G., and De Maria, R. (2016). Epithelial-mesenchymal transition: a new target in anticancer drug discovery. Nat Rev Drug Discov *15*, 311-325.

Maréchal, A., and Zou, L. (2013). DNA Damage Sensing by the ATM and ATR Kinases. Cold Spring Harbor Perspectives in Biology *5*.

Markiewicz, A., Topa, J., Nagel, A., Skokowski, J., Seroczynska, B., Stokowy, T., Welnicka-Jaskiewicz, M., and Zaczek, A.J. (2019). Spectrum of Epithelial-Mesenchymal Transition Phenotypes in Circulating Tumour Cells from Early Breast Cancer Patients. Cancers (Basel) *11*.

Marnett, L.J., and Plastaras, J.P. (2001). Endogenous DNA damage and mutation. Trends in Genetics *17*, 214-221.

Masuda, T.A., Inoue, H., Sonoda, H., Mine, S., Yoshikawa, Y., Nakayama, K., Nakayama, K., and Mori, M. (2002). Clinical and biological significance of S-phase kinase-associated protein 2 (Skp2) gene expression in gastric carcinoma: modulation of malignant phenotype by Skp2 overexpression, possibly via p27 proteolysis. Cancer Res *62*, 3819-3825.

McCormack, M.P., Young, L.F., Vasudevan, S., de Graaf, C.A., Codrington, R., Rabbitts, T.H., Jane, S.M., and Curtis, D.J. (2010). The Lmo2 Oncogene Initiates Leukemia in Mice by Inducing Thymocyte Self-Renewal. *327*, 879-883.

Mechetner, E., Kyshtoobayeva, A., Zonis, S., Kim, H., Stroup, R., Garcia, R., Parker, R.J., and Fruehauf, J.P. (1998). Levels of multidrug resistance (MDR1) P-glycoprotein expression by human breast cancer correlate with in vitro resistance to taxol and doxorubicin. Clinical Cancer Research *4*, 389-398.

Meidhof, S., Brabletz, S., Lehmann, W., Preca, B.T., Mock, K., Ruh, M., Schuler, J., Berthold, M., Weber, A., Burk, U.*, et al.* (2015). ZEB1-associated drug resistance in cancer cells is reversed by the class I HDAC inhibitor mocetinostat. EMBO Mol Med *7*, 831-847.

Mejlvang, J., Kriajevska, M., Vandewalle, C., Chernova, T., Sayan, A.E., Berx, G., Mellon, J.K., and Tulchinsky, E. (2007). Direct repression of cyclin D1 by SIP1 attenuates cell cycle progression in cells undergoing an epithelial mesenchymal transition. Mol Biol Cell *18*, 4615-4624.

Miquelajauregui, A., Van de Putte, T., Polyakov, A., Nityanandam, A., Boppana, S., Seuntjens, E., Karabinos, A., Higashi, Y., Huylebroeck, D., and Tarabykin, V. (2007). Smad-interacting protein-1 (Zfhx1b) acts upstream of Wnt signaling in the mouse hippocampus and controls its formation. Proceedings of the National Academy of Sciences *104*, 12919-12924.

Mironchik, Y., Winnard, P.T., Jr., Vesuna, F., Kato, Y., Wildes, F., Pathak, A.P., Kominsky, S., Artemov, D., Bhujwalla, Z., Van Diest, P.*, et al.* (2005). Twist overexpression induces in vivo angiogenesis and correlates with chromosomal instability in breast cancer. Cancer Res *65*, 10801-10809.

Modrich, P. (2006). Mechanisms in Eukaryotic Mismatch Repair. Journal of Biological Chemistry *281*, 30305-30309.

Morel, A.-P., Ginestier, C., Pommier, R.M., Cabaud, O., Ruiz, E., Wicinski, J., Devouassoux-Shisheboran, M., Combaret, V., Finetti, P., Chassot, C.*, et al.* (2017). A stemness-related ZEB1–MSRB3 axis governs cellular pliancy and breast cancer genome stability. Nature Medicine *23*, 568.

Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S., and Puisieux, A. (2008). Generation of Breast Cancer Stem Cells through Epithelial-Mesenchymal Transition. PLOS ONE *3*, e2888.

Mowat, D.R., Croaker, G.D., Cass, D.T., Kerr, B.A., Chaitow, J., Adès, L.C., Chia, N.L., and Wilson, M.J. (1998). Hirschsprung disease, microcephaly, mental retardation, and characteristic facial features: delineation of a new syndrome and identification of a locus at chromosome 2q22-q23. Journal of Medical Genetics *35*, 617-623.

Mowat, D.R., Wilson, M.J., and Goossens, M. (2003). Mowat-Wilson syndrome. Journal of Medical Genetics *40*, 305-310.

Musgrove, E.A., Davison, E.A., and Ormandy, C.J. (2004). Role of the CDK inhibitor p27 (Kip1) in mammary development and carcinogenesis: insights from knockout mice. J Mammary Gland Biol Neoplasia *9*, 55-66.

Narod, S.A., and Foulkes, W.D. (2004). BRCA1 and BRCA2: 1994 and beyond. Nature Reviews Cancer *4*, 665.

Nieto, M.A. (2002). The snail superfamily of zinc-finger transcription factors. Nat Rev Mol Cell Biol *3*, 155-166.

Nieto, M.A., Huang, Ruby Y.-J., Jackson, Rebecca A., and Thiery, Jean P. (2016). EMT: 2016. Cell *166*, 21-45.

O'Driscoll, M., and Jeggo, P.A. (2006). The role of double-strand break repair — insights from human genetics. Nature Reviews Genetics *7*, 45.

Oesterle, E.C., Chien, W.-M., Campbell, S., Nellimarla, P., and Fero, M.L. (2011). p27(Kip1) is required to maintain proliferative quiescence in the adult cochlea and pituitary. Cell Cycle *10*, 1237-1248.

Omilusik, K.D., Best, J.A., Yu, B., Goossens, S., Weidemann, A., Nguyen, J.V., Seuntjens, E., Stryjewska, A., Zweier, C., Roychoudhuri, R.*, et al.* (2015). Transcriptional repressor ZEB2 promotes terminal differentiation of CD8⁺ effector and memory T cell populations during infection. *212*, 2027-2039.

Pal, T., Permuth-Wey, J., Betts, J.A., Krischer, J.P., Fiorica, J., Arango, H., LaPolla, J., Hoffman, M., Martino, M.A., Wakeley, K.*, et al.* (2005). BRCA1 and BRCA2 mutations account for a large proportion of ovarian carcinoma cases. Cancer *104*, 2807-2816.

Pascal, J.M., O'Brien, P.J., Tomkinson, A.E., and Ellenberger, T. (2004). Human DNA ligase I completely encircles and partially unwinds nicked DNA. Nature *432*, 473-478.

Paul-Konietzko, K., Thomale, J., Arakawa, H., and Iliakis, G. (2015). DNA Ligases I and III Support Nucleotide Excision Repair in DT40 Cells with Similar Efficiency. Photochem Photobiol *91*, 1173-1180.

Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nat Rev Cancer *7*, 415-428.

Postigo, A.A. (2003). Opposing functions of ZEB proteins in the regulation of the TGFβ/BMP signaling pathway. The EMBO Journal *22*, 2443-2452.

Postigo, A.A., and Dean, D.C. (1999). ZEB represses transcription through interaction with the corepressor CtBP. Proceedings of the National Academy of Sciences of the United States of America *96*, 6683-6688.

Postigo, A.A., Depp, J.L., Taylor, J.J., and Kroll, K.L. (2003). Regulation of Smad signaling through a differential recruitment of coactivators and corepressors by ZEB proteins. The EMBO journal *22*, 2453-2462.

Puisieux, A., Brabletz, T., and Caramel, J. (2014). Oncogenic roles of EMT-inducing transcription factors. Nat Cell Biol *16*, 488-494.

Qin, Y., Capaldo, C., Gumbiner, B.M., and Macara, I.G. (2005). The mammalian Scribble polarity protein regulates epithelial cell adhesion and migration through E-cadherin. J Cell Biol *171*, 1061-1071.

Reisman, D., Glaros, S., and Thompson, E.A. (2009). The SWI/SNF complex and cancer. Oncogene *28*, 1653-1668.

Remacle, J.E., Kraft, H., Lerchner, W., Wuytens, G., Collart, C., Verschueren, K., Smith, J.C., and Huylebroeck, D. (1999). New mode of DNA binding of multi-zinc finger transcription factors: deltaEF1 family members bind with two hands to two target sites. The EMBO journal *18*, 5073-5084.

Rieder, C.L., and Palazzo, R.E. (1992). Colcemid and the mitotic cycle. J Cell Sci *102 (Pt 3)*, 387-392.

Roth, D.B. (2014). V(D)J Recombination: Mechanism, Errors, and Fidelity. Microbiology Spectrum *2*.

Sanchez-Tillo, E., Lazaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E.C., Castells, A., Engel, P., and Postigo, A. (2010). ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. Oncogene *29*, 3490-3500.

Saxena, M., Stephens, M.A., Pathak, H., and Rangarajan, A. (2011). Transcription factors that mediate epithelial-mesenchymal transition lead to multidrug resistance by upregulating ABC transporters. Cell Death Dis *2*, e179.

Sayan, A.E., Griffiths, T.R., Pal, R., Browne, G.J., Ruddick, A., Yagci, T., Edwards, R., Mayer, N.J., Qazi, H., Goyal, S.*, et al.* (2009). SIP1 protein protects cells from DNA damage-induced apoptosis and has independent prognostic value in bladder cancer. Proc Natl Acad Sci U S A *106*, 14884-14889.

Schmidt, Johanna M., Panzilius, E., Bartsch, Harald S., Irmler, M., Beckers, J., Kari, V., Linnemann, Jelena R., Dragoi, D., Hirschi, B., Kloos, Uwe J.*, et al.* (2015). Stem-Cell-like Properties and Epithelial Plasticity Arise as Stable Traits after Transient Twist1 Activation. Cell Reports *10*, 131-139.

Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc *3*, 1101-1108.

Schneider-Poetsch, T., Ju, J., Eyler, D.E., Dang, Y., Bhat, S., Merrick, W.C., Green, R., Shen, B., and Liu, J.O. (2010). Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. Nat Chem Biol *6*, 209-217.

Seyfried, T.N., and Huysentruyt, L.C. (2013). On the Origin of Cancer Metastasis. Critical reviews in oncogenesis *18*, 43-73.

Sharma, S.S., and Pledger, W.J. (2016). The non-canonical functions of p27(Kip1) in normal and tumor biology. Cell Cycle *15*, 1189-1201.

Shioiri, M., Shida, T., Koda, K., Oda, K., Seike, K., Nishimura, M., Takano, S., and Miyazaki, M. (2006). Slug expression is an independent prognostic parameter for poor survival in colorectal carcinoma patients. British Journal Of Cancer *94*, 1816.

Sicinski, P., Zacharek, S., and Kim, C. (2007). Duality of p27Kip1 function in tumorigenesis. Genes Dev *21*, 1703-1706.

Siebzehnrubl, F.A., Silver, D.J., Tugertimur, B., Deleyrolle, L.P., Siebzehnrubl, D., Sarkisian, M.R., Devers, K.G., Yachnis, A.T., Kupper, M.D., Neal, D.*, et al.* (2013). The ZEB1 pathway links glioblastoma initiation, invasion and chemoresistance. *5*, 1196-1212.

Singh, D.K., Krishna, S., Chandra, S., Shameem, M., Deshmukh, A.L., and Banerjee, D. (2014). Human DNA ligases: a comprehensive new look for cancer therapy. Med Res Rev *34*, 567-595.

Slingerland, J., and Pagano, M. (2000). Regulation of the Cdk inhibitor p27 and its deregulation in cancer. Journal of Cellular Physiology *183*, 10-17.

Soni, R., O'Reilly, T., Furet, P., Muller, L., Stephan, C., Zumstein-Mecker, S., Fretz, H., Fabbro, D., and Chaudhuri, B. (2001). Selective in vivo and in vitro effects of a small molecule inhibitor of cyclin-dependent kinase 4. J Natl Cancer Inst *93*, 436-446.

Soza, S., Leva, V., Vago, R., Ferrari, G., Mazzini, G., Biamonti, G., and Montecucco, A. (2009). DNA ligase I deficiency leads to replication-dependent DNA damage and impacts cell morphology without blocking cell cycle progression. Mol Cell Biol *29*, 2032-2041.

Stephens, P.J., Tarpey, P.S., Davies, H., Loo, P.V., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R.*, et al.* (2012). The landscape of cancer genes and mutational processes in breast cancer. Nature *486*, 400-404.

Takagi, T., Moribe, H., Kondoh, H., and Higashi, Y. (1998). DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages. Development *125*, 21-31.

Talmadge, J.E., and Fidler, I.J. (2010). AACR centennial series: the biology of cancer metastasis: historical perspective. Cancer Res *70*, 5649-5669.

Tam, W.L., and Weinberg, R.A. (2013). The epigenetics of epithelial-mesenchymal plasticity in cancer. Nat Med *19*, 1438-1449.

Taylor, A.M., Groom, A., and Byrd, P.J. (2004). Ataxia-telangiectasia-like disorder (ATLD)-its clinical presentation and molecular basis. DNA Repair (Amst) *3*, 1219-1225.

Theunissen, J.W.F., and Petrini, J.H.J. (2006). Methods for Studying the Cellular Response to DNA Damage: Influence of the Mre11 Complex on Chromosome Metabolism. In Methods in Enzymology (Academic Press), pp. 251-284.

Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871-890.

Thompson, S.L., and Compton, D.A. (2008). Examining the link between chromosomal instability and aneuploidy in human cells. The Journal of Cell Biology *180*, 665-672.

Tigli, H., Buyru, N., and Dalay, N. (2005). Molecular analysis of the p27/kip1 gene in breast cancer. Mol Diagn *9*, 17-21.

Tong, J., Fu, Y., Xu, X., Fan, S., Sun, H., Liang, Y., Xu, K., Yuan, Z., and Ge, Y. (2014). TGF-β1 Stimulates Human Tenon's Capsule Fibroblast Proliferation by miR-200b and its Targeting of p27/kip1 and RND3. Investigative Ophthalmology & Visual Science *55*, 2747-2756.

Van de Putte, T., Francis, A., Nelles, L., van Grunsven, L.A., and Huylebroeck, D. (2007). Neural crest-specific removal of Zfhx1b in mouse leads to a wide range of neurocristopathies reminiscent of Mowat–Wilson syndrome. Human Molecular Genetics *16*, 1423-1436.

Van de Putte, T., Maruhashi, M., Francis, A., Nelles, L., Kondoh, H., Huylebroeck, D., and Higashi, Y. (2003). Mice Lacking Zfhx1b, the Gene That Codes for Smad-Interacting Protein-1, Reveal a Role for Multiple Neural Crest Cell Defects in the Etiology of Hirschsprung Disease–Mental Retardation Syndrome. The American Journal of Human Genetics *72*, 465-470.

van Helden, M.J., Goossens, S., Daussy, C., Mathieu, A.-L., Faure, F., Marçais, A., Vandamme, N., Farla, N., Mayol, K., Viel, S.*, et al.* (2015). Terminal NK cell maturation is controlled by concerted actions of T-bet and Zeb2 and is essential for melanoma rejection. *212*, 2015-2025.

van Roy, F., and Berx, G. (2008). The cell-cell adhesion molecule E-cadherin. Cellular and Molecular Life Sciences *65*, 3756-3788.

Vandewalle, C., Van Roy, F., and Berx, G. (2009). The role of the ZEB family of transcription factors in development and disease. Cellular and Molecular Life Sciences *66*, 773-787.

Vasen, H.F.A., Blanco, I., Aktan-Collan, K., Gopie, J.P., Alonso, A., Aretz, S., Bernstein, I., Bertario, L., Burn, J., Capella, G.*, et al.* (2013). Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut *62*, 812-823.

Vega, S., Morales, A.V., Ocana, O.H., Valdes, F., Fabregat, I., and Nieto, M.A. (2004). Snail blocks the cell cycle and confers resistance to cell death. Genes Dev *18*, 1131-1143.

Verschueren, K., Remacle, J.E., Collart, C., Kraft, H., Baker, B.S., Tylzanowski, P., Nelles, L., Wuytens, G., Su, M.-T., Bodmer, R.*, et al.* (1999). SIP1, a Novel Zinc Finger/Homeodomain Repressor, Interacts with Smad Proteins and Binds to 5′-CACCT Sequences in Candidate Target Genes. Journal of Biological Chemistry *274*, 20489-20498.

Vincent, T., Neve, E.P., Johnson, J.R., Kukalev, A., Rojo, F., Albanell, J., Pietras, K., Virtanen, I., Philipson, L., Leopold, P.L.*, et al.* (2009). A SNAIL1-SMAD3/4 transcriptional repressor complex promotes TGF-beta mediated epithelial-mesenchymal transition. Nat Cell Biol *11*, 943-950.

Wakamatsu, N., Yamada, Y., Yamada, K., Ono, T., Nomura, N., Taniguchi, H., Kitoh, H., Mutoh, N., Yamanaka, T., Mushiake, K.*, et al.* (2001). Mutations in SIP1, encoding Smad interacting protein-1, cause a form of Hirschsprung disease. Nature Genetics *27*, 369.

Wallace, S.S. (2014). Base excision repair: A critical player in many games. DNA Repair *19*, 14-26.

Wang, J., Scully, K., Zhu, X., Cai, L., Zhang, J., Prefontaine, G.G., Krones, A., Ohgi, K.A., Zhu, P., Garcia-Bassets, I.*, et al.* (2007). Opposing LSD1 complexes function in developmental gene activation and repression programmes. Nature *446*, 882.

Webster, A.D.B., Barnes, D.E., Lindahl, T., Arlett, C.F., and Lehmann, A.R. (1992). Growth retardation and immunodeficiency in a patient with mutations in the DNA ligase I gene. The Lancet *339*, 1508-1509.

Wellner, U., Schubert, J., Burk, U.C., Schmalhofer, O., Zhu, F., Sonntag, A., Waldvogel, B., Vannier, C., Darling, D., zur Hausen, A.*, et al.* (2009). The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. Nat Cell Biol *11*, 1487-1495.

Wiechelman, K.J., Braun, R.D., and Fitzpatrick, J.D. (1988). Investigation of the bicinchoninic acid protein assay: identification of the groups responsible for color formation. Anal Biochem *175*, 231-237.

Witsch, E., Sela, M., and Yarden, Y. (2010). Roles for growth factors in cancer progression. Physiology (Bethesda) *25*, 85-101.

Yamada, K., Yamada, Y., Nomura, N., Miura, K., Wakako, R., Hayakawa, C., Matsumoto, A., Kumagai, T., Yoshimura, I., Miyazaki, S.*, et al.* (2001). Nonsense and frameshift mutations in ZFHX1B, encoding Smad-interacting protein 1, cause a complex developmental disorder with a great variety of clinical features. Am J Hum Genet *69*, 1178-1185.

Yang, M.-H., Hsu, D.S.-S., Wang, H.-W., Wang, H.-J., Lan, H.-Y., Yang, W.-H., Huang, C.-H., Kao, S.-Y., Tzeng, C.-H., Tai, S.-K.*, et al.* (2010). Bmi1 is essential in Twist1-induced epithelial–mesenchymal transition. Nature Cell Biology *12*, 982.

Yao, G. (2014). Modelling mammalian cellular quiescence. Interface Focus *4*, 20130074.

Yokoi, S., Yasui, K., Saito-Ohara, F., Koshikawa, K., Iizasa, T., Fujisawa, T., Terasaki, T., Horii, A., Takahashi, T., Hirohashi, S.*, et al.* (2002). A Novel Target Gene, SKP2, within the 5p13 Amplicon That Is Frequently Detected in Small Cell Lung Cancers. The American Journal of Pathology *161*, 207-216.

Yoon, D.-S., Wersto, R.P., Zhou, W., Chrest, F.J., Garrett, E.S., Kwon, T.K., and Gabrielson, E. (2002). Variable Levels of Chromosomal Instability and Mitotic Spindle Checkpoint Defects in Breast Cancer. The American Journal of Pathology *161*, 391-397.

Yu, M., Bardia, A., Wittner, B.S., Stott, S.L., Smas, M.E., Ting, D.T., Isakoff, S.J., Ciciliano, J.C., Wells, M.N., Shah, A.M.*, et al.* (2013). Circulating Breast Tumor Cells Exhibit Dynamic Changes in Epithelial and Mesenchymal Composition. *339*, 580-584.

Yuen, H.-F., Chan, Y.-P., Lok-Yee Wong, M., Kwok, W.-K., Chan, K.-K., Lee, P.-Y., Srivastava, G., Law, S.Y.-K., Wong, Y.-C., Wang, X.*, et al.* (2007). Upregulation of Twist in oesophageal squamous cell carcinoma is associated with neoplastic transformation and distant metastasis. Journal of Clinical Pathology *60*, 510-514.

Zahreddine, H., and Borden, K. (2013). Mechanisms and insights into drug resistance in cancer. Frontiers in Pharmacology *4*.

Zhang, J., Wang, P., Wu, F., Li, M., Sharon, D., Ingham, R.J., Hitt, M., McMullen, T.P., and Lai, R. (2012). Aberrant expression of the transcriptional factor Twist1 promotes invasiveness in ALK-positive anaplastic large cell lymphoma. Cellular Signalling *24*, 852-858.

Zhang, P., Sun, Y., and Ma, L. (2015). ZEB1: at the crossroads of epithelial-mesenchymal transition, metastasis and therapy resistance. Cell Cycle *14*, 481-487.

Zhang, P., Wang, L., Rodriguez-Aguayo, C., Yuan, Y., Debeb, B.G., Chen, D., Sun, Y., You, M.J., Liu, Y., Dean, D.C.*, et al.* (2014a). miR-205 acts as a tumour radiosensitizer by targeting ZEB1 and Ubc13. Nat Commun *5*, 5671.

Zhang, P., Wei, Y., Wang, L., Debeb, B.G., Yuan, Y., Zhang, J., Yuan, J., Wang, M., Chen, D., Sun, Y.*, et al.* (2014b). ATM-mediated stabilization of ZEB1 promotes DNA damage response and radioresistance through CHK1. Nat Cell Biol *16*, 864-875.

Zhao, G.Y., Sonoda, E., Barber, L.J., Oka, H., Murakawa, Y., Yamada, K., Ikura, T., Wang, X., Kobayashi, M., Yamamoto, K.*, et al.* (2007). A critical role for the ubiquitin-conjugating enzyme Ubc13 in initiating homologous recombination. Mol Cell *25*, 663-675.