Novel Targets and Strategies in the Treatment of Liver and Oral Pharyngeal Cancer

Rhys Gillman, BBiomedSc

For the degree of Master of Philosophy

College of Public Health, Medical and Veterinary Sciences

James Cook University

March 2021

Acknowledgements

Throughout the preparation and execution of this thesis, I have received a great deal of support from a number of individuals.

I would first like to thank my supervisor, Associate Professor Lionel Hebbard, for his unwavering support. Lionel provided me with the freedom to follow my instincts and to learn from my mistakes, but also the structure that I needed to flourish with my research. Additionally, I would like to acknowledge Lionel's wife, Doctor Miriam Wankell, with whom I spent countless hours in the laboratory and learned, from her expertise, lessons that will remain with me throughout my career.

I wish also to acknowledge the rest of my advisory team, Associate Professor Patrick Schaeffer and Doctor Ira Cooke. Both of you brought not only the specific expertise which allowed certain components of this thesis to exist, but also at times the friendly ear that I needed to work my way out of a rut.

My time throughout this degree was made not only tolerable, but truly a joyful experience, by the presence of my colleagues Lauren Taylor, Brittany Dewdney, and Casey Toft, all of whom I can now count as close friends. I am truly privileged to enjoy the presence of such talented and enthusiastic individuals over morning coffee.

Importantly, I wish to thank my family for being the foundation on which I was able to pursue this degree. Firstly, is my partner in both life and research, Rebecca Bourquin, who provided not only the care and support I needed at home, but also at times the support I needed in the laboratory. Next, are my parents and siblings, who never fail to produce a consistent sense of acceptance and understanding for my pursuits.

Thank you as well to the clinician Doctor Kylie Lopes Floro. While I only met you towards the end of my journey through this project, you made a significant impact on my readiness to complete this piece of work with both your intellectual and physical input into the work.

I also would like to thank the various associations, departments and institutions at the university who provide the facilities and student support for not only myself, but for my many colleagues. This includes the Graduate Research School, the JCU and Molecular and Cell Biology Student Associations, and the College of Public Health, Medical and Veterinary Sciences administration team.

Finally, I wish to acknowledge those who provided the funding for this project, both in the form of grant funds and a student stipend scholarship. This includes the Cancer Council, James Cook University, and the Australian Government.

Thesis Declaration

I, Rhys Gillman, declare that the work presented in this thesis has been completed for the degree of Master of Philosophy under the supervision of Associate Professor Lionel Hebbard, Associate Professor Patrick Schaeffer, and Doctor Ira Cooke.

This thesis has not been submitted or accepted for any other degree or professional qualification in any university or tertiary institution, except as specified for partial aspects of the work.

This thesis was written entirely on my own and does not contain any material previously written by another person. The work produced in this thesis is original and was completed by myself, or is a result of a substantial contribution to the projects, in which case this contribution has been clearly indicated in the text and is explicitly stated in the statement of contribution of others (next page). This thesis contains work that has been published which was solely written by myself.

This work is in no way a violation or infringement of copyright, trademark, patent, or any other rights of any person.

Signature:

Date:

Statement of the Contribution of Others

This thesis contains work that has been contributed to by others. The specific contributions of others are listed below:

Chapter 1.3.2

Preliminary Data – Identification of HORMAD2 as a Novel HCC Gene

Contributions: Preliminary data in Figure 1.1 was provided by a previous student of the Hebbard Lab Mohammed Alanzay.

Chapter 1.3.4

The Role of DNA Damage and Repair in Liver Cancer (Published Review)

Authors: Rhys Gillman, Kylie Lopes Floro, Miriam Wankell, Lionel Hebbard

Contributions: The candidate (**RG**) conceptualised and conducted the literature search, wrote the original draft, and conducted revision and editing of the manuscript. KF, MW and LH revised the draft and suggested additions to the manuscript.

Chapter 4

***In vitro* Characterisation of HORMAD2 in HCC**

Contributions: The work presented in this chapter was a result of a joint contribution with PhD Candidate Brittany Dewdney. The candidate (**RG**) conceptualised and performed all experimentation with BD with equal contributions. BD generated the knock-down lentiviral constructs, while **RG** generated the over-expression lentiviral constructs.

This chapter contains *in vitro* data material that will be used in the thesis of Brittany Dewdney for the degree of Doctor of Philosophy submitted to James Cook University (2021), which has been written as a separate and original work to this thesis.

Chapter 5

Molecular Pathways of Smoke-Induced OPC Pathogenesis

Contributions: As outlined at the beginning of Chapter 5, this chapter represents my contribution of a collaborative project with Doctor Kylie Lopes Floro. Experimental treatment of oral cancer cells and preparation of samples for sequencing was performed entirely by KF, LH, and MW. The candidate (**RG**) performed all data preparation and bioinformatic analyses of the data. The candidate (**RG**) also reviewed the literature and wrote the introduction and discussion of this data. KF and LH revised the draft and provided feedback.

List of Abbreviations

AGRF	Australian Genome Research Facility
ALD	Alcoholic liver disease
ALDH	Alcohol Dehydrogenase
aNHEJ	Alternative Non-Homologous End-Joining
APN	Adiponectin
ATP	Adenosine Tri-Phosphate
BER	Base-Excision Repair
BMI	Body Mass Index
BP	Biological Process
CDS	Coding Sequence
CO	Crossover
CSC	Cancer Stem Cell
DDR	DNA Damage Response
DE	Differential Expression
DEN	Diethylnitrosamine
DNA	Deoxyribonucleic Acid
DSB	Double Stranded Break
DSBR	Double Stranded Break Repair
DSF-GTP	Differential Scanning Fluorometry of a GFP-Tagged Protein
ECM	Extracellular matrix
EMT	Epithelial-Mesenchymal Transition
FBS	Foetal Bovine Serum
GDAC	Genome Data Analysis Centre
GFP	Green-Fluorescent Protein
GFP-BASTA	GFP-Based Stability Assay
GFP-EMSA	GFP-Based Electrophoretic Mobility Shift Assay
GGR	Global Genomic Repair
GO	Gene Ontology
GORILLA	Gene Ontology enRIchment anaLysis and visuaLizAtion
HBV	Hepatitis B Virus
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C Virus
HH	High HORMAD2
HIS	Histidine Tag
HPV	Human papilloma virus
HR	Homologous Recombination
LH	Low HORMAD2
LIHC	Liver Hepatocellular Carcinoma
LOH	Loss of Heterozygosity
LPS	Lipopolysaccharide
miRNA	Micro-RNA
MMP	Metalloprotease
MMR	Mismatch Repair
MSCI	Meiotic Sex Chromosome Inactivation
MSUC	Meiotic Silencing of Unsynapsed Chromsomes
NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non-Alcoholic Steatohepatitis
NER	Nucleotide Excision Repair
NH	Normal HORMAD2
NO	Nitric Oxide
nt	Nuclotide
OPC	Oral-Pharyngeal Cancer
PCA	Prinicple Component Analysis
PE	Plating Efficiency
PIKK	Phosphoinositide 3-Kinase-Related Kinase
QC	Quality Control
REVIGO	Reduce and Visualise Gene Ontology
RNASeq	Ribonucleic Acid Sequencing
RNS	Reactive Nitrogen Species
ROS	Reactive Oxygen Species
SF	Surviving Fraction
shRNA	Small Hairpin RNA
SNP	Single Nucleotide Polymorphism

ssDNA	Single Stranded DNA
TCGA	The Cancer Genome Atlas
TCR	Transcription-Coupled Repair
TLS	Translesion Synthesis
UCSC	University of California Santa Cruz
UTR	Untranslated Region
VEGF	Vascular Endothelial Growth Factor

Abstract

Hepatocellular carcinoma (HCC) and oral-pharyngeal cancer (OPC) together constitute 13.5% of global cancer related mortality. Firstly, previous work in the Hebbard lab identified HORMAD2, a poorly-studied protein involved in sex cell division, as having a potentially novel role in HCC. To elucidate this role, an *in silico* bioinformatic analysis of HCC sequencing data was performed, and HORMAD2 function was studied *in vitro* using HCC cell lines and recombinant protein expressed in *E. coli*. The *in silico* results suggest that HORMAD2 acts a tumour suppressor which responds to oxidative stress in the liver by down-regulating the cell cycle, and HORMAD2 expression is consistently lost in HCC, which may play a role in its pathogenesis. Furthermore, *in vitro* data indicates that the knockdown of HORMAD2 in HCC promotes rapid repair of DNA, and may promote cancer stem cell formation or insulin-mediated cell proliferation. Secondly, based on observations in the Hebbard lab of increased invasiveness of OPC cells exposed to cigarette smoke and radiation, a bioinformatic analysis was performed to identify pathways responsible. This investigation revealed that smoke and radiation act synergistically to enhance the invasiveness of OPC, which may explain the increased occurrence of metastases in patients undergoing radiotherapy.

Table of Contents

List of Figures

List of Tables

List of Appendices

Chapter 1: Introduction

1.1 Preamble

A section of this introduction was published in January 2021 in Biochimica et Biophysica Acta (BBA) - Reviews on Cancer under the title **"The role of DNA damage and repair in liver cancer"** (DOI: 10.1016/j.bbcan.2020.188493). The introduction and final chapters of the original publication have been altered for the purpose of cohesion in this thesis, along with all chapter, figure and table formatting.

1.2 Cancer in Australia

In simple terms, cancer can be defined by the key capabilities neoplastic cells acquire which ultimately allow for chronic, unregulated proliferation. These capabilities, known as the hallmarks of cancer, are acquired successively, driven primarily by genomic instability which allows for the selective introduction of genetic alterations (1). Cancers are a major contributor to the burden of disease, both in Australia and globally. Indeed, in the context of decades of improved understanding, detection, and treatment of numerous cancers, a few select forms of cancer remain not only prominent, but show increasing mortality rates. For the purpose of this research piece, these cancers include liver cancer, and oral-pharyngeal cancer (2).

1.3 Liver Cancer

Primary liver cancer (PLC) is the second most common cause of cancer-related death worldwide (3). While the mortality rates associated with most other cancers have decreased over the last two decades, liver cancer mortality continues to rise (4). The two common subtypes of PLC are hepatocellular carcinoma (HCC) and cholangiocarcinoma, which make up approximately 75% and 12% of cases respectively (5). The major risk factors for HCC include hepatitis B and C viral infection (HBV and HCV, respectively), alcoholic liver disease (ALD), non-alcoholic fatty liver disease (NAFLD) and its progressive form non-alcoholic steatohepatitis (NASH)(6). Primary liver disease often arises in the context of inflammation and fibrosis; the excessive accumulation of extracellular matrix proteins, that can result in cirrhosis and cancer (7, 8). In 2015, HCC accounted for over 800 thousand deaths globally (9), a figure which was higher than the three decades prior, and is expected to continue rising due to the increasing prevalence of obesity and obesity-related diseases, a driver of HCC formation and progression (10). Thus, the burden of HCC is a significant worldwide public health problem.

A major contributing factor to HCC mortality is the lack of available treatment options. Most common chemotherapeutic agents like cisplatin, which work by inducing DNA damage in rapidly replicating cells, are ineffective for HCC treatment. The only effective chemotherapy drug available for HCC is

Sorafenib, which extends survival rates by a dismal 2-3 months (11). Thus, HCC is a major global health concern.

1.3.1 The Role of Fructose and Adiponectin in HCC

The introduction of improved antiviral drugs for viral hepatitis has initiated a shift in the primary aetiology of HCC. Between the years of 2006 to 2014, HCV as a cause of HCC dropped from approximately 65% of cases to less than 50%. Simultaneously, NASH-related HCC increased from approximately 8% of cases to almost 20% (12). Thus, while cases of viral-related HCC have decreased, the incidence of fatty liver-associated HCC is growing steadily and is expected to become the primary cause of HCC in coming years (12, 13). Additionally, NAFLD-associated HCC has been shown to present significantly lower survival rates than HCV-associated HCC (14). Thus, NAFLD-associated HCC is a major and increasing concern, yet its pathogenesis remains poorly understood and treatment options remain limited.

NAFLD and NASH are the liver manifestation of the metabolic syndrome and are associated with obesity. Indeed, clear associations have been demonstrated between body-mass-index (BMI) and liver cancer risk (15). It is generally understood that the pathogenesis behind this is low-grade inflammation. Obesity-related inflammation occurs due to an imbalance in the secretion of adipokines (of which leptin and adiponectin have the most well documented role in NALFD) and inflammatory cytokines tumour necrosis factor alpha (TNFα) and interleukin 6 (IL-6) by adipose tissue itself. Specifically, during the development of NAFLD the adipocyte secretion of anti-inflammatory adiponectin decreases, while pro-inflammatory leptin, IL-6 and TNFα secretion increases (16, 17), causing an overall pro-inflammatory state. This inflammation exacerbates reactive oxygen species (ROS) production, which causes DNA damage in the form of single-base modifications and double-stranded breaks. The resulting genomic instability drives cancer-promoting mutagenesis.

Recently, the intake of fructose has become an intense area of focus in the metabolic syndrome and in NAFLD (18-21). Fructose is a simple carbohydrate. The use of sweeteners like high-fructose corn syrup (HFCS) in foods such as soft-drinks and condiments has led to a dramatic increase in fructose consumption in the western diet. In contrast to glucose, fructose is taken up and metabolised by hepatocytes independently of rate-limiting, regulatory steps such as glucokinase and phosphofructokinase. Thus, excess fructose consumption leads to over-production of acetyl-CoA and subsequent fat production (22), therefore increasing the burden of obesity. To further its negative effects, fructose consumption has also been shown to induce inflammation and exacerbate ROS production (19), exacerbating DNA damage.

1.3.2 Preliminary Data – Identification of HORMAD2 as a novel HCC Gene

In light of the growing concern for NAFLD-associated HCC, and the potential involvement of adiponectin (APN) and fructose in its pathogenesis, the Hebbard lab experimentally investigated these

factors. They demonstrated that APN-knockout (APN-KO) mice treated with the chemical carcinogen diethylnitrosamine (DEN) produce tumours 7-fold larger than wild-type controls (23). Subsequently, in unpublished experiments, APN-KO mice were fed a high-fructose diet. Unexpectedly, these mice presented with 31-fold smaller tumours than APN-KO mice fed a normal chow (**Figure 1.1A**). Interestingly, the decreased tumour growth was associated with the down-regulation of a poorly-studied gene called *HORMAD2* (**Figure 1.1B**). HORMAD2 protein expression was further found in several HCC cell lines, while others did not express the protein (**Figure 1.1C**) (data not published).

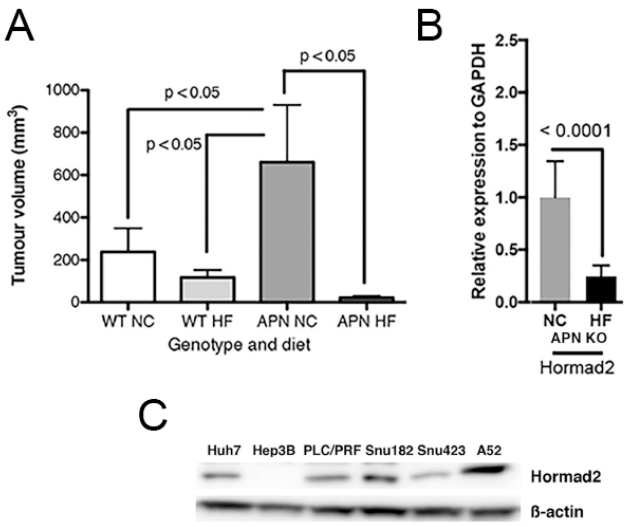

Figure 1.1: HORMAD2 expression in HCC. A) APN knockout mice present with large tumours, but a high fructose (HF) diet reduced tumour growth 31-fold (all $p<0.05$, One-Way ANOVA and Newman-Keuls multiple comparison test), associated with **B)** decreased *HORMAD2* gene expression ($p < 0.0001$ Mann Whitney) (Expression relative to GAPDH and control). **C)** HORMAD2 protein is expressed in several HCC cell lines (data not published). APN, Adiponectin; HF, high-fructose diet; KO, knock-out; NC, normal chow; WT, wild-type.

1.3.3 HORMAD2

HORMA (Hop1/Rev7/Mad2) domain containing 2 (HORMAD2) is a poorly understood protein, known to play a role in chromosomal synapsis surveillance during meiosis, the specialised process of cell division which results in the production of sex cells. As such, the expression of HORMAD2 is typically restricted to the testes. Structurally, HORMAD2 is a 307 amino acid (aa) nucleoprotein, predicted to contain a single N-terminal HORMA domain and a C-terminal disordered tail lacking inherent structure.

1.3.3.1 HORMA Domain Proteins

The HORMA domain was first identified by Aravind in 1998 (24). It was suggested to function as a recognition domain of chromatin states that result from DNA adducts, double-stranded breaks, or non-attachment to the spindle apparatus, as well as an adaptor domain, which mediates interactions between proteins within networks of cell-cycle control and DNA repair. The HORMA domain is named after the three yeast (*Saccharomyces cerevisiae*) proteins that share the domain: **Ho**p1, **R**ev7p, and **Ma**d2. HORMA domain-containing proteins have since been identified in a range of species across many

domains of life, including nematode (*C. elegans*), yeast (*S. cerevisiae*), human, and bacterial. In all cases, the HORMA domain itself is highly conserved.

A common feature of all HORMA domain proteins is their association with chromatin. Upon activation, HORMA domain proteins form hierarchical structures through interactions of their HORMA domains with closure motifs on their partner, mediating both self-recruitment and recruitment of each other to chromosomes (25). These interactions involve large conformational changes that have only been demonstrated in Mad2 and Hop1 (26).

A number of HORMA domain-containing proteins exist in the human proteome (**Figure 1.2**). Rev7 is a component of DNA polymerase zeta involved in translesion DNA synthesis. Mad2 and P31^{comet} both function in the spindle-assembly checkpoint through inhibition of CDC20 (27). Atg13 and Atg101 function in the autophagy pathway. Finally, HORMAD1 and HORMAD2 are the two meiotic mammalian HORMADs.

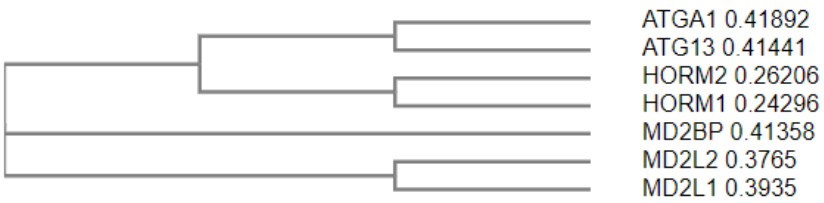

Figure 1.2: Phylogenetic tree of human HORMA domain-containing proteins. Multiple sequence alignment and tree generation was performed using Clustal Omega and Simple Phylogeny. MD2BP, P31^{comet}; MD2L2, REV7; MD2L1, MAD2

1.3.3.2 Function

All of HORMAD2's known functions come from studies in meiotic cells and its involvement in the pachytene checkpoint of meiosis. Meiosis is a specialised form of cell division resulting in the formation of haploid gametes, and occurs only within the sex organs. Synapsis is a crucial component of meiosis which promotes genetic diversity and the accurate segregation of chromosomes in meiotic synapsis relies on physical connections between homologous chromosomes, known as chiasmata, which are formed following genetic crossover (CO) events. CO involves the formation of double-stranded breaks (DSBs) by the meiotic recombination protein SPO11, exonucleolytic end resection, strand invasion promoted by the meiotic RAD51 homolog DNA meiotic recombinase 1 (DMC1), and repair by replication with the template (28), a process analogous with homologous recombination repair in somatic cells.

Successful synapsis is monitored by the pachytene checkpoint, a surveillance system whose molecular nature is still highly debated but is related to the DNA damage response. The pachytene checkpoint removes male or female sex cells (meiocytes) with asynaptic chromosomes which have failed to form genetic COs. One way this is achieved is through transcriptional silencing of asynaptic chromosomes,

known as meiotic silencing of unsynapsed chromosomes (MSUC), which occurs due to the accumulation of γH2AX heterochromatin (29). This process is prominent in male sex cells between the intrinsically asynaptic X and Y sex chromosomes (30), specifically termed meiotic sex chromosome inactivation (MSCI). Functionally, MSUC silences transcription of essential genes and leads to elimination of the cell. Similarly, though with opposite consequences, MSCI prevents transcription of otherwise toxic genes on the sex chromosomes. Failed synapsis can also trigger the formation of spontaneous DSBs in a currently unknown manner, which triggers the checkpoint kinase 2 (CHEK2)-mediated DNA damage response and apoptosis. In these ways, asynaptic meiocytes are removed (31).

HORMAD2 localises at unsynapsed chromosomes and is required for MSUC/MSCI (**Figure 1.3**) (32-34). Indeed, male HORMAD2-knock-out (KO) mice are infertile and females are not (33, 34). This is because the failure of MSCI causes apoptosis in male spermatocytes, while in females, the failure of MSUC prevents apoptosis in oocytes with asynaptic chromosomes. Thus, while this has not been investigated, it could be postulated that females with HORMAD2 deficiency may have a higher likelihood of producing offspring with genetic diseases related to aneuploidy.

HORMAD2 is recruited to asynaptic chromosomes in a HORMAD1-dependent manner (33), mediated by closure-motif interactions with the HORMA domain (25). HORMAD2 and BRCA1 are required for accumulation of ATR at unsynapsed chromosomes (32-35), though the nature of interactions that take place are unknown. It is debated whether HORMAD2 is required for the recruitment of other DDR factors including BRCA1 and TOPBP1. Nonetheless, after its HORMAD2-dependent recruitment, ATR is responsible for the H2AX phosphorylation which produces MSUC/MSCI, while ATM may also participate (35, 36). Following successful synapsis and formation of the synaptonemal complex (SC), HORMAD2 is depleted from chromosomes in a TRIP13-dependent manner (32).

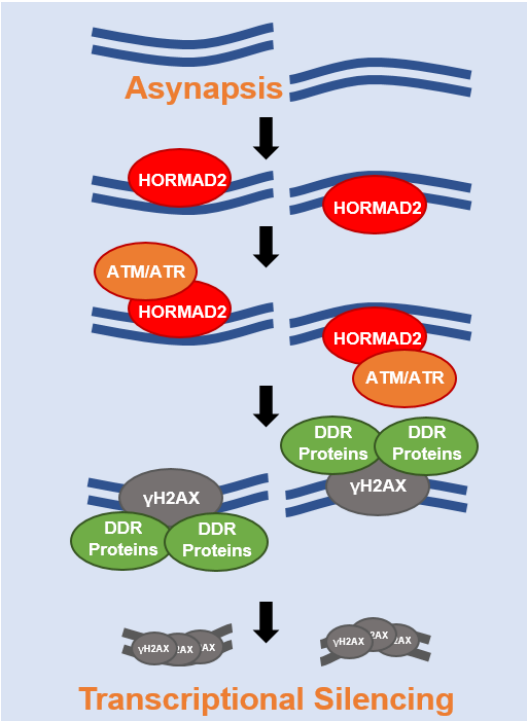

Figure 1.3: Postulated roles of HORMAD2 in MSUC/MSCI. Experimental evidence shows that HORMAD2 is localized at unsynapsed chromosomes and required for the recruitment of ATR. ATR phosphorylates H2AX, forming γH2AX. DDR proteins (BRCA1, 53BP1, TOPBP1) localize at γH2AX. Heterochromatin spreads across the entire chromosome and it is silenced.

Conformational changes may be required to mediate protein-protein interactions with HORMAD2. Such conformational changes could be induced by binding to chromatin or DNA, interactions with other proteins, or via modifications such as phosphorylation. DNA-binding by HORMAD2 is unstudied, and therefore this possibility needs to be evaluated. Additionally, the closure-motif-HORMA domain interaction between HORMAD1 and HORMAD2 is a conserved feature of highly-related yeast protein Mad2, which exhibits conformational changes during interaction (25), and thus, this interaction could mediate other protein interactions. Finally, Fukuda, Pratto *et al.* (36) demonstrated that BRCA1 promoted phosphorylated forms of HORMAD1 and HORMAD2. Reduced HORMAD2 phosphorylation was associated with decreased recruitment of MSUC proteins to unsynapsed chromosomes. Thus, any of these interactions or modifications could induce conformational changes required for further protein-protein interactions with HORMAD2.

Another possible function of HORMAD2 is in inter-homolog (IH) bias (**Figure 1.4**). Due to cohesin-dependent proximity, the sister chromatid is the intrinsically preferred template for recombination repair in somatic cells. In meiosis, however, this bias must be overcome by generating a barrier to sister chromatid recombination (BSCR), which produces IH bias (37-39). This usually involves inhibition of homologous recombination repair (HRR) with the sister chromatid. Indeed, this is supported by the fact that HORMAD2 is depleted from chromosome axes immediately after synapsis, as after this point it is

beneficial to repair the remaining DSBs in any way possible. Thus, if expressed in somatic cells, HORMAD2 may inhibit recombination with the sister chromatid and impede HRR. Many HORMAD2 orthologs are required for IH bias in other organisms, including Hop1 in yeast and Him-3 in *C. elegans* (32).

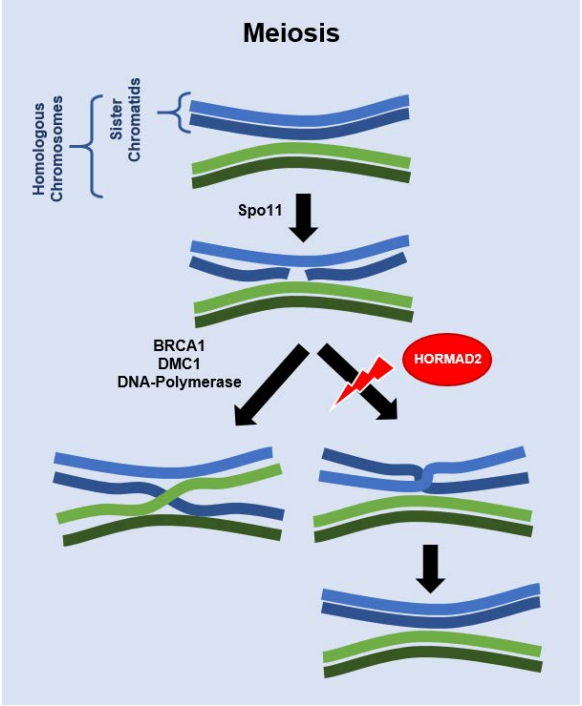

Figure 1.4: Alternative suggested function of HORMAD2 in IH bias. The primary difference between meiotic and mitotic recombination is the partner choice for recombination. In meiosis, recombination with the homologous chromosome is required to mediate accurate chromosome segregation and increase genetic diversity. HORMAD2 is suggested to contribute to IH bias by inhibiting repair with the sister chromatid.

It should be noted that since the original study by Kogo, Tsutsumi *et al.* (33), other studies have showed little or no expression of HORMAD2 in the ovaries (40). Considering then that accurate synapsis and correct chromosomal segregation is no more important in male physiology than it is in female physiology, this raises the question as to whether the proposed function of HORMAD2 is correct. It is here conjectured that a possible explanation for this sex disparity in HORMAD2 expression is one of natural selection pressure. Because of its proposed role in MSCI, HORMAD2 deficiency in a male is understood to render an individual incapable of producing fertile offspring, possibly due to expression of cytotoxic signals from the asynaptic sex chromosomes. In females however, a deficiency in HORMAD2 only impairs the ovary's ability to remove oocytes via MSUC. Thus, it is conceivable, however unsupported by experimental evidence, that this impaired MSUC could increase the number of viable oocytes and therefore increase fertility. Were this the case, this scenario would produce much higher evolutionary pressure in males to maintain higher HORMAD2 expression.

1.3.3.3 Structure

No experimentally derived 3-dimensional structure exists for HORMAD2. However, Rosenberg described the structure of the HORMA domain in detail (41). Briefly, the HORMA domain is usually approximately 200aa long, and consists of 2 distinct regions, a 150aa core followed by a 50aa safety belt. The core consists of 3 alpha helices, a 3-stranded beta sheet and hairpin region of 2 additional beta strands (**Figure 1.5**). Like HORMAD2, some proteins, such as Mad2 and Rev7 are composed almost entirely of the HORMA domain, while others, like Hop1, are attached to other functional domains like DNA-binding Zn finger domains.

In HORMA domain proteins, homomeric and/or hetermeric interactions are mediated by interaction of the closure motif with the safety belt of the HORMA domain. HORMAD2 has been shown to contain a closure motif within its disordered C-terminal tail (position 283-307) which is capable of interacting with HORMAD1 and likely with the HORMAD2 HORMA domain itself (25).

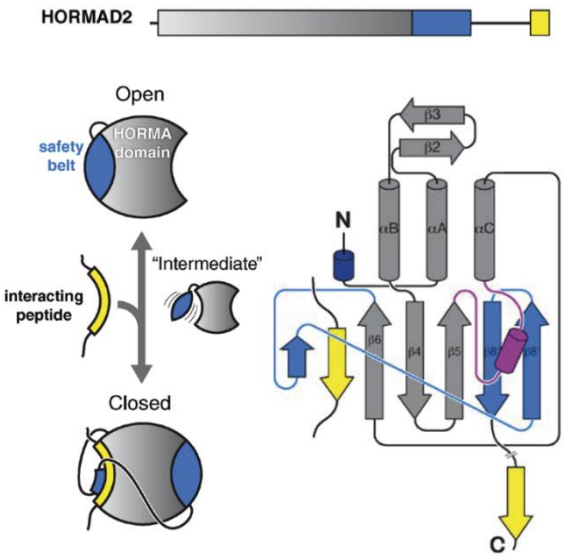

Figure 1.5: Proposed structure of the HORMAD2 HORMA domain by Rosenberg and Corbett (41). Blue = safety belt, Yellow = interacting peptides (Closure Motif), Purple = unique loop structure

1.3.3.4 HORMADs in Cancer

HORMAD1 and HORMAD2 have both been identified as cancer-testis antigens (CTAs) (40). This means that their expression is normally restricted to the testes, but they often become aberrantly expressed in cancers. CTAs are commonly implicated in genomic stability (42). Indeed, considering the involvement of HORMA-domain containing proteins in meiotic DNA recombination (41), it is likely that ectopic HORMAD2 expression would influence cancer genomic stability. However, based on meiotic studies of HORMAD2, it remains unclear whether its expression would have a positive or negative impact on genomic stability. On one hand, the role on HORMAD2 in recruiting DDR proteins to chromosomes suggests that it could promote DNA repair. However, its postulated role in BSCR, which involves inhibition of HRR, suggests the opposite. Further complicating their potential roles in

cancer is the fact that while the loss of HRR can drive genomic instability by negatively impacting repair of DNA damage, the over-expression of HRR enzymes can also drive genomic instability by increasing the rate of large DNA structural alterations (43). The potential roles of HORMAD2 in somatic cell cancer are outlined in the figure below (**Figure 1.6**).

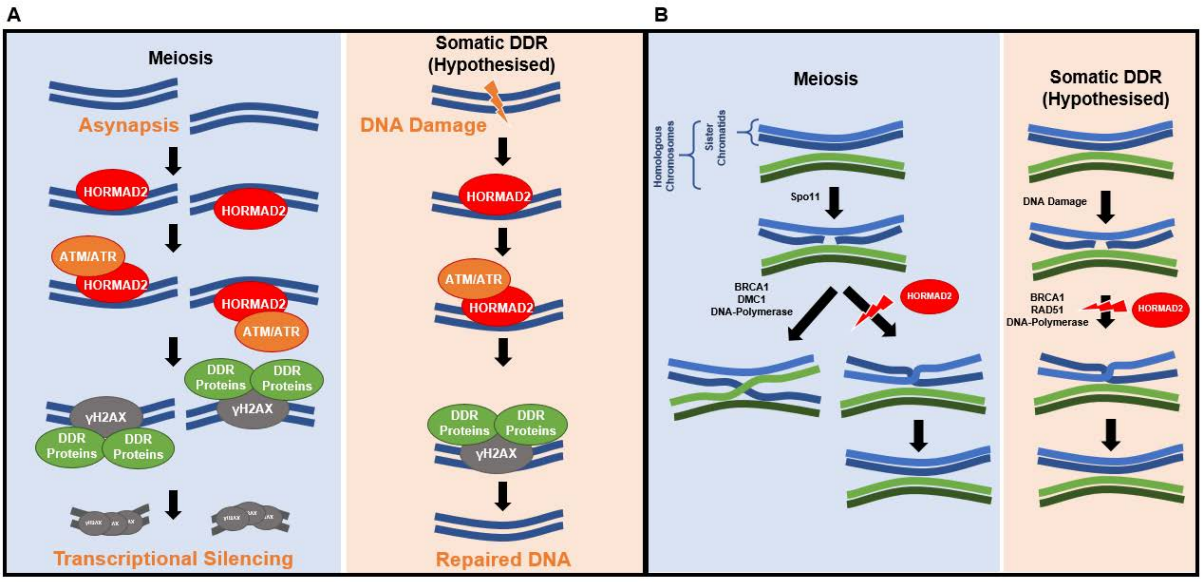

Figure 1.6: Postulated functions of HORMAD2 in somatic cells and cancer. A) HORMAD2 is required for ATR recruitment in MSUC/MSCI, which results in γH2AX accumulation and recruitment of other DDR proteins. Therefore, in somatic cells, HORMAD2 may perform a similar role by recruiting ATR and other DDR proteins to sites of DNA damage to facilitate repair. Alternatively, B) HORMAD2 may be important for BSCR

Reports have shown that HORMAD2 is over-expressed in the tumours of lung cancer patients, with localised expression in the cell nuclei as observed in meiocytes (40). Conversely, HORMAD2 is repressed in thyroid cancer due to hypermethylation, correlating with poor prognosis (44). Additionally, HORMAD2 single nucleotide polymorphisms (SNPs) can affect survival in advanced non-small-cell lung cancer (45). Thus, while HORMAD2 appears to play an important role in cancer, its function remains unclear.

In contrast, the highly related HORMAD2 homolog, HORMAD1, has been extensively studied in cancer, though its mechanistic role is also unclear. HORMAD1 is over-expressed in 80% of basal-like breast cancers, and is associated with reduced sensitivity to PARP inhibitors (46). Two independent studies have also found HORMAD1 over-expressed in lung cancer and localised to the nucleus in response to DNA damage (47, 48). However, one study concluded that HORMAD1 inhibited HRR activity (47), while the other concluded that HRR activity was increased and that HORMAD1 depletion enhanced sensitivity to PARP inhibitors (48).

1.3.3.5 HORMAD2 in the Liver

Multiple reports indicate that HORMAD2 is expressed in the normal liver (40, 49) **(Figure 1.7)**. Decreasing HORMAD2 gene expression in the liver has been identified as a marker of NAFLD progression (50), suggesting that it could play a role in NAFLD progression. However, no further research has been performed to elucidate the function of HORMAD2 in the liver.

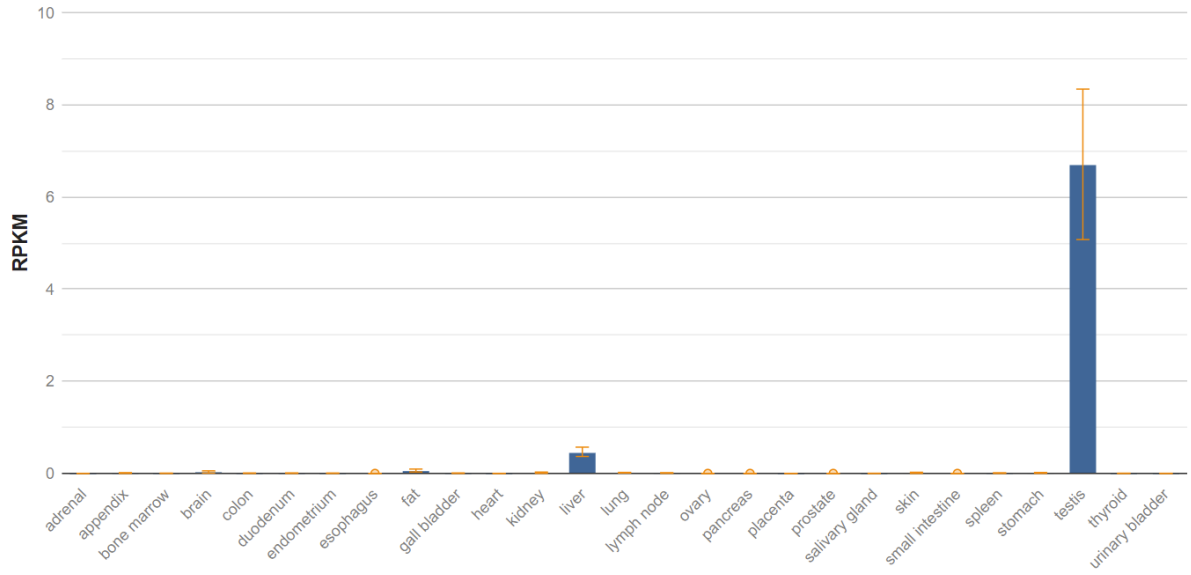

Figure 1.7: Mean RNA expression of HORMAD2 in 27 different normal tissues. HORMAD2 expression is detected in the healthy tissue of both testis and liver. Figure retrieved from the National Centre for Biotechnology Information. Expression was evaluated via RNA-seq from 95 human individuals in a publication by the Human Protein Atlas (HPA) (49). RPKM, reads per kilobase per million reads.

1.3.4 The Role of DNA Damage and Repair in Liver Cancer (Published Review)

Rhys Gillman[1], Kylie Lopes Floro[1,3], Miriam Wankell[1,4], Lionel Hebbard[1,2, 4]

[1]Department of Molecular and Cell Biology, College of Public Health, Medical, and Veterinary Sciences, James Cook University, Townsville, Queensland, Australia

[2]Storr Liver Centre, Westmead Institute for Medical Research, Westmead Hospital and University of Sydney, Sydney, New South Wales, Australia.

[3]Department of Radiation Oncology, Townsville University Hospital, Townsville, Queensland, Australia

[4]Australian Institute for Tropical Health and Medicine, Townsville, Queensland, Australia

Chronic liver diseases represent a major worldwide health burden and contribute to the development of primary liver cancer, of which hepatocellular carcinoma is the most common (5). The major risk factors for HCC include hepatitis B and C viral infection alcoholic liver disease, non-alcoholic fatty liver disease and its progressive form non-alcoholic steatohepatitis (6). Primary liver disease often arises in the context of inflammation and fibrosis; the excessive accumulation of extracellular matrix proteins, that can result in cirrhosis and cancer (7, 8). In 2010, HCC accounted for almost 1 million deaths (51), which is suspected to be an underestimation (52). These figures were higher than the three decades prior, and are expected to continue rising, due to the increasing prevalence of obesity and obesity-related diseases, a driver of HCC formation and progression (10). Thus, the burden of HCC is a significant worldwide public health problem.

Chronic liver diseases of all aetiologies are associated with genomic instability, and these changes are detectable long before dysplasia occurs (53). Furthermore, people who develop HCC tend to have higher sensitivity to DNA damage (54), suggesting that genomic instability plays a role in the development of liver cancer. Genomic stability is monitored by the DNA damage response (DDR), which includes DNA repair pathways and cell-cycle checkpoints which stall the replication of damaged cells. The five major DNA repair pathways which maintain genomic stability in eukaryotes are base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), homologous recombination (HR), and non-homologous end-joining (NHEJ). The distinction between these repair pathways is often ignored in cancer research, likely due to the broad nature of DNA damage involved in cancer progression.

DNA damage and its repair are a double-edged sword to cancer development. On the one hand, the breakdown of DNA repair, which increases genomic instability and mutagenesis, is a fundamental hallmark of cancer (1). Mutagenesis allows the development of gain-of-function mutations in oncogenes, or loss-of-function mutations in tumour-suppressor genes, both of which can drive neoplastic growth. On the other hand, uncontrolled mutagenesis and breakdown of genomic stability can be detrimental to tumour cells due to the mutation of crucial genes, stalling of replication forks, and/or triggering of apoptosis.

Given that genomic instability can be detrimental, cancer cells frequently evolve mechanisms to increase their capacity for DNA repair, allowing them to become tolerant to some level of DNA damage. Importantly, this means that cancers which arise as a consequence of deficient DNA repair subsequently become dependent on other redundant DNA repair pathways for survival. In recent years, the targeting of these cancer-dependent pathways has demonstrated efficacy in cancer therapy, in a concept known as 'synthetic lethality' (55).

The purpose of this review was to analyse the literature on DNA repair pathways in liver diseases, and to investigate the possible roles of these pathways in the development of HCC. Genome-wide association studies have identified many potential genetic variants associated with liver cancer risk, some of which are in DNA repair enzymes, and some of these variants have been further investigated in observational and experimental studies. Moreover, this review aims to discuss the importance of each DNA repair pathway in liver cancer, and how this information may assist in the development of cancer therapies. Furthermore, for the purpose of this review, the types of DNA damage and their repair will be primarily discussed in isolation, however it should be remembered that repair pathways are often redundant and capable of considerable cross-talk (56).

1.3.4.1 Single base damage and base excision repair

A common feature of all major risk factors for liver cancer development is inflammation, which leads to oxidative DNA damage. Oxidative damage and other small modifications to the nitrogenous bases of DNA are predominantly repaired via the base-excision repair (BER) pathway (**Figure 1.8**). Here, the damaged base is excised by specific glycosylases such as human 8-Oxoguanine glycosylase (hOgg1) or Nei-Like DNA glycosylase 1 (NEIL1), yielding a single apurinic/apyrimidinic (AP) site. AP endonuclease 1 (APE1) then recognises the AP site and makes an incision 5' of the AP site, resulting in a 1 nucleotide (nt) gap with a bound deoxyribose phosphate (dRP) 3' of the excision. DNA polymerase beta (POLβ) then fills the gap and removes the 5' dRP through 5' dRP-lyase activity. In some cases, modification of the 5'dRP requires extra processing over a 2-11 nt region by the flap endonuclease (FEN1) and POLδ/ε, known as long-patch repair. Finally, BER ends with repair of the nick by DNA ligase III with its co-factor X-ray Cross-complementing protein 1 (XRCC1). This pathway has been discussed extensively in previous reviews (57, 58).

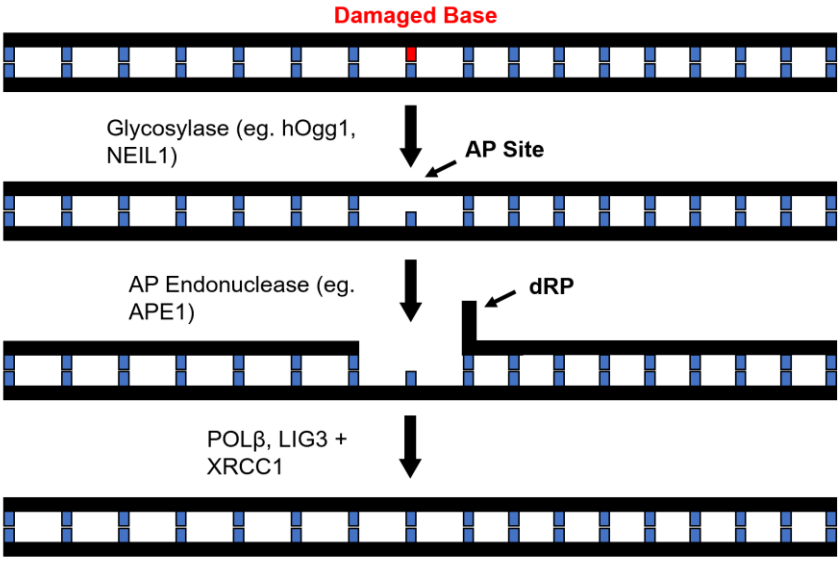

Figure 1.8: Repair of DNA base damage through base-excision repair (BER). BER repairs DNA damage through excision of the damaged base by specific glycosylase enzymes, removing the apurinic/apyrimidinic (AP) site by AP endonuclease, and replacing the missing base by DNA polymerase beta (POLβ), using the complementary DNA strand as reference. See text for details.

1.3.4.1.1 Sources of single base damage in the liver

Acting essentially as a biological filter between the digestive tract and the rest of the body (59), the healthy liver is subjected to oxidative DNA damage on a daily basis. Indeed, 80% of the liver's blood supply comes directly from the gut via the portal vein, carrying with it viable gut-associated bacteria, bacterial products and environmental toxins that have crossed the gut membrane (60). Hence, this makes the liver the first point of contact with inflammatory microbial molecules like lipopolysaccharide (LPS), to which the immune system mounts an inflammatory response resulting in the production of reactive oxygen species (ROS) (61, 62). Due to the hepatic clearance of the toxin, LPS is usually present in the portal vein at a much higher concentration than the peripheral circulation, meaning the liver is exposed to higher LPS levels than other organs (63). Inflammation is also associated with the production of nitric oxide (NO) and reactive nitrogen species (RNS) by inducible nitric oxide synthase (iNOS), further exacerbating oxidative DNA damage (64). In addition to these exogenous sources, liver hepatocytes have abundant mitochondria, which produce endogenous ROS during oxidative phosphorylation. Thus, any impediment in the reductive capacity of hepatocytes could lead to *in situ* increases in ROS.

In this light, pathologies that affect the liver have the concomitant action of further increasing inflammation and oxidative damage. For example, NAFLD and NASH generally occur with obesity, which induces chronic inflammation through a shift in cytokine secretion from the adipose tissue, wherein expression of anti-inflammatory adiponectin is reduced while pro-inflammatory molecules like interleukin-6 (IL-6) and tumour necrosis factor alpha (TNFα) increase (16, 17). Further, obesity-related cholesterol overload and aberrant lipogenesis may play an important role in HCC development (65), and it has been demonstrated that a high cholesterol diet in diethylnitrosamine (DEN) treated mice

causes increased oxidative damage (66). Indeed, convincing epidemiological associations have been observed between body-mass index (BMI) and liver cancer risk (15). Alcohol contributes to oxidative damage in the liver by the production of ROS through increased cytochrome P450s activity (67), and by increasing LPS levels in the portal vein through increased gut permeability (68-70). In addition, chronic HBV/HCV infection is another source of immune-related inflammation, and much of the associated DNA damage is mediated by reactive nitrogen due to iNOS activation (71, 72). Together, patients with both obesity and viral hepatitis experience a much higher risk of HCC (73), suggesting that these sources of oxidative damage can act cumulatively. Finally, these sources of liver damage trigger a positive feedback system wherein resident Kupffer cells secrete further inflammatory cytokines to further amplify the inflammatory response (74).

1.3.4.1.2 Modulation of base excision repair in HCC

Given the prominence of oxidative damage in the liver, BER may be important for HCC pathogenesis, however few studies have experimentally investigated this. Some studies have focussed on hOgg1, the glycosylase that removes modified guanine, 8-oxo-7,8-dihydroguanine (8-oxo-dG), the most common result of oxidative DNA damage (75). Jaiswal, LaRusso *et al.* (64) demonstrated that NO is able to directly inhibit BER by interacting with hOgg1 to form S-nitrosothiol adducts (64). As NO is produced during inflammation, this causes an enhanced effect of oxidative damage and may contribute to HCC mutagenesis. In support, Kakehashi, Ishii *et al.* (76) found that homozygous mouse Ogg1 deletion promoted progression from hepatocellular adenoma to HCC after treatment with phenobarbital, a compound which promotes HCC via increased ROS production. Furthermore, knockout of another glycosylase, NEIL1, has been shown to induce broad features of metabolic syndrome, including NAFLD (77), suggesting that oxidative DNA damage may play an early role in the development of NAFLD. Moreover, the importance of BER in HCC was reinforced in findings by Di Maso, Mediavilla *et al.* (78), demonstrating that APE1 is up-regulated in human HCC.

While external to BER, enzymes responsible for directly preventing oxidative damage to DNA may also influence HCC pathogenesis. The human mutT homologs (MTH1 and MTH2) are examples of such enzymes, which hydrolyse oxidised nucleotides to prevent their introduction into DNA. Lin, Liu *et al.* (79) recently demonstrated an effect of HBV in reducing expression of MTH1 and MTH2 via the viral HBV X protein (HBx), resulting in markedly increased 8-oxo-dG levels. Further, a recent bioinformatics study by Yu, Wang *et al.* (80) listed the up-regulation of MTH1 as a powerful predictor of HCC.

These results show that enzymes responsible for dealing with oxidative DNA damage become up-regulated in HCC, indicating the presence and importance of oxidative damage. Further, the results of several studies show that when these enzymes are removed, either experimentally or by the actions of

reactive molecules and viral proteins, HCC risk increases due to an inability to correct oxidative DNA damage.

1.3.4.1.3 Variants in base excision repair may mediate HCC susceptibility

If BER is important to liver cancer progression, single nucleotide polymorphisms (SNPs) in BER enzymes could influence cancer predisposition. The literature investigating associations between BER SNPs and HCC is summarised (**Table 1.1**). By example, the presence of a cysteine allele rather than serine at hOGG1 codon 326 (Ser326Cys) has been reported to significantly increase the risk of HCC in a study of Chinese patients with high rates of HBV infection (81). In this same study, a variant at codon 280 of XRCC1 (Arg280His) did not significantly affect HCC risk, however patients with both mutations had higher HCC risk again, suggesting that the effect of these variants could be cumulative within a pathway. Other polymorphisms in XRCC1 have also been investigated in relation to HCC. For example, in other Chinese studies with prevalent HBV infection, XRCC1 Arg399Gln and Arg194Tryp caused a slight increase in HCC risk (82, 83), while a Taiwanese study found the effect insignificant (84). Contrastingly, in Egyptian, HCV-infected individuals, XRCC1 Arg399Gln and Arg280His were observed to be significantly protective for HCC, while Arg194Tryp increased HCC risk. This Egyptian population also appeared to have vastly different allelic frequencies in these genes to the other studies. Clearly, these inconsistent effects need to be investigated further and may provide insight into differing vulnerabilities to HBV and HCV-associated HCC.

Table 1.1: Polymorphisms in DNA repair enzymes studied in association with hepatocellular carcinoma. For each polymorphism, data is given for the risk (OR) attributed to an individual carrying a single risk allele (heterozygous), both risk alleles (homozygous) or the risk associated with carrying either one or two risk alleles. The aetiology/risk factors column gives approximate indication of the proportions of each risk factor for the study cohorts. **Note:** Where studies have investigated multiple polymorphisms, sample sizes are listed for each polymorphism. Thus, adding up sample sizes across polymorphisms is not representative of the total individuals studied. BER, Base-excision repair; HBV, Hepatitis B virus; HCV, Hepatitis C virus; HR, Homologous recombination; MMR, Mismatch repair; NER, Nucleotide excision repair; NHEJ, Non-homologous end-joining; OR, Odds-Ratio

Pathway	Gene	Polymorphism	Single Allele OR	Double Allele OR	Any Risk Allele OR	Study Population	n (Cases)	n (Controls)	Year	Aetiology/Risk Factors	Reference
BER	hOGG1	Ser326Cys	2.31	2.94	2.38	Chinese	350	400	2009-2011	80% HBV, 78% Alcohol, 67% Smoker	(81)
BER	XRCC1	Arg280His	1.38	0.91	1.33	Chinese	350	400	2009-2011	80% HBV, 78% Alcohol, 67% Smoker	(81)
BER	XRCC1	Arg280His	0.53	1.09	NR	Egyptian	87	88	Not reported	100% HCV	(85)
BER	XRCC1	Arg399Gln	1.42	1.4	NR	Chinese	202	236	2008-2010	52% HBV, 12% HCV, 63% Alcohol, 44% Smoker	(82)
BER	XRCC1	Arg399Gln	0.72	0.75	NR	Egyptian	87	88	Not reported	100% HCV	(85)
BER	XRCC1	Arg399Gln	1.16	1.74	1.5	Chinese	410	410	2008-2011	36% HBV, 5.1% HCV, 41% Alcohol, 37% Smoker	(83)
BER	XRCC1	Arg399Gln	1.1	1.57	NR	Taiwanese	577	389	1997-2001	100% HBV	(84)
BER	XRCC1	Arg194Tryp	2.14	NR	NR	Egyptian	87	88	Not reported	100% HCV	(85)
BER	XRCC1	Arg194Trp	1.17	2.26	1.42	Chinese	410	410	2008-2011	36% HBV, 5.1% HCV, 41% Alcohol, 37% Smoker	(83)
NER	XPD	Asp312Asn	1.12	0.89	NR	Chinese	712	635	2006-2008	72% HBV, 18% HCV	(86)
NER	XPD	Asp312Asn	1.23	1.66	1.37	Chinese	410	410	2008-2011	36% HBV, 5.1% HCV, 41% Alcohol, 37% Smoker	(83)
NER	XPD	Lys751Gln	1.75	2.47	NR	Chinese	712	635	2006-2008	72% HBV, 18% HCV	(86)
NER	XPD	Lys751Gln	1.14	3.51	1.42	Chinese	410	410	2008-2011	36% HBV, 5.1% HCV, 41% Alcohol, 37% Smoker	(83)
NER	XPD	Lys751Gln	0.91	0.23	NR	Taiwanese	577	389	1997-2001	100% HBV	(84)
HR/MMR	EXO1	Glu589Lys	0.90	2.15	1.08	Turkish	224	224	1005-2011	59.4% HBV, 24% HCV, 28.6% Alcohol, 47.8% Smoker	(87)
HR	NBS1	Gln185Glu	1.41	2.27	NR	Chinese	865	900	2006-2011	83% HBV, 47.1% Alcohol, 59% Smoker	(88)
HR	NBS1	Gln185Glu	NR	NR	1.19	Chinese	481	581	2014-2016	100% HBV, 61% Alcohol, 64% Smoker	(89)
HR	NBS1	rs2735383 (3' UTR)	0.93	1.09	NR	Chinese	865	900	2006-2011	83% HBV, 47.1% Alcohol, 59% Smoker	(88)
HR	NBS1	rs10464867 (3' UTR)	NR	NR	1.16	Chinese	481	581	2014-2016	100% HBV, 61% Alcohol, 64% Smoker	(89)
HR	NBS1	rs1063053 (3' UTR)	NR	NR	0.89	Chinese	481	581	2014-2016	100% HBV, 61% Alcohol, 64% Smoker	(89)
HR	NBS1	Asp399Asp	NR	NR	1.12	Chinese	481	581	2014-2016	100% HBV, 61% Alcohol, 64% Smoker	(89)
HR	RAD52	rs7963551 (3' UTR)	0.88	0.65	NR	Chinese	1806	1954	2009-2012	100% HBV, 65% Alcohol, 61% Smoker	(90)
NHEJ	XRCC7 (DNA-PKcs)	rs7003908 (Intron)	0.88	0.66	0.85	Taiwanese	298	889	2004-2010	69% Alcohol, 75% Smoker	(91)
MMR	MSH2	rs2303428 (Intron)	1.76	1.85	1.82	Chinese	1021	1021	2009-2015	83% HBV, 33% Alcohol, 34% Smoker	(92)
MMR	MLH1	rs1800734 (5' UTR)	1.53	1.85	NR	Taiwanese	577	389	1997-2001	100% HBV	(84)
MMR	MLH1	rs1800734 (5' UTR)	1.22	1.75	1.29	Chinese	1036	1036	2009-2015	83% HBV, 33% Alcohol, 34% Smoker,	(93)

1.3.4.2 Bulky DNA lesions and repair by nucleotide excision or translesion synthesis

Intrinsically linked to the digestive system, the liver inevitably suffers exposure to external toxins, many of which are known risk factors to HCC development and cause DNA damage in the form of bulky lesions which distort the DNA double helix. These types of lesions are repaired via the nucleotide excision repair (NER) pathway, which was reviewed in depth by Spivak (94) (**Figure 1.9**). Briefly, two sub-pathways of NER exist, transcription-coupled repair (TCR) and global genomic repair (GGR), which differ only in their initial recognition of the lesion. In TCR, RNA polymerase II (RNAPII) becomes stalled at a lesion during transcription, resulting in recruitment of the Cockayne syndrome group A and B (CSA and CSB) protein complexes. RNAPII is then removed and the protein complex Transcription Factor II H (TFIIH) recruited to the site. Conversely, GGR occurs independently of transcription. In GGR, lesions are recognised either by the first of the Xeroderma Pigmentosum complementation proteins, group C (XPC) protein, in the case of large-distorting lesions, or they are initially recognised by the XP group E (XPE) complex which provides the helix distortion required for recognition by XPC. The XPC complex then opens the local region of DNA, allowing recruitment of TFIIH to the damaged DNA region. After TFIIH recruitment the pathways converge and dual 5' and 3' incisions are created on either side of the lesion. To achieve this, TFIIH components XPB and XPD form a 20-30nt bubble around the lesion, allowing recruitment of other factors including XPA, XPG, and XPF to the complex with excision repair cross-complementing group 1 (ERCC1). ERCC1-XPF then performs the 5' incision, while XPG performs the 3' incision and DNA polymerase (epsilon (ε) in replicating cells, delta (δ) and kappa (κ) in non-replicating cells) fills the single-stranded DNA (ssDNA) gap. Finally, the remaining nick is repaired by DNA ligase I (LIG1) in replicating cells, or through DNA ligase III (LIG3)-XRCC1 activity in non-replicating cells.

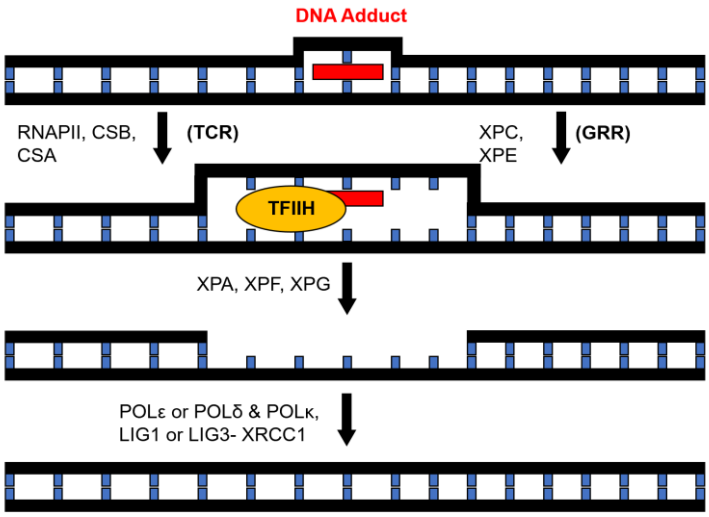

Figure 1.9: Repair of DNA adducts via nucleotide excision repair (NER). Larger DNA lesions require excision of a considerable region of surrounding bases. This is achieved by either transcription-coupled repair (TCR) or global genomic repair (GGR). TCR and GGR differ in their initial recognition process of DNA adducts, but converge in their mechanisms for removal of surrounding DNA bases, followed by template-directed repair. See text for details.

Should a lesion escape repair by NER, cells have adapted an alternate mechanism of replication known as translesion synthesis (TLS), which allows cells to replicate DNA despite the presence of a bulky lesion. In contrast to high fidelity polymerases, which normally replicate DNA faithfully, TLS Y-family polymerases eta (η), iota (ι), kappa (κ), and Rev1, and B-family polymerase zeta (ζ) contain more flexible and therefore erroneous active sites. This allows bulky lesions to be bypassed, introducing extra random nucleotides on the newly synthesised strand in the process (95). Each member of the TLS polymerases are adapted to deal specifically with different types of lesions (96), and can also bypass oxidative lesions (97). The recruitment of TLS polymerases is mediated by proliferating cell nuclear antigen (PCNA), and promoted by the RAD6/RAD18-mediated monoubiquitination of PCNA (98).

1.3.4.2.1 Sources of bulky DNA lesions in the liver DNA

Exposure to aflatoxin B1 (AFB1) is a major risk factor in HCC development in sub-Saharan Africa and South-East China, where 80% of HCC cases occur (8). AFB1 is a toxin produced by *Aspergillus flavus and parasiticus* fungus which commonly contaminates improperly stored foods. The genotoxicity of AFB1 relies on biotransformation by cytochrome P450, producing the reactive metabolite AFB1-8,9-epoxide, which may spontaneously form AP sites repaired through BER, or persistent DNA adducts (99). Additionally, chemotherapeutics to which HCC is notoriously resistant such as cisplatin work by inducing DNA adducts. These persistent adducts are repaired primarily by NER, or bypassed by TLS polymerases.

1.3.4.2.2 Nucleotide excision repair deficiencies can drive HCC

Disruption of the NER pathway and overactivity of the TLS pathway likely promotes AFB1-induced HCC development. With respect to NER disruption, it has been shown that XPA is required for the efficient removal of AFB1 adducts in human fibroblasts (100). Furthermore, mice with homozygous knockout of XPA have increased susceptibility to AFB1-induced HCC (101). The TLS pathway may also influence this type of HCC susceptibility. A common feature of AFB1-induced HCC is a G to T transversion at codon 249 of the tumour-suppressor p53, which occurs in approximately 50% of HCCs from areas with high levels of AFB1 contamination (102, 103). AFB1 specifically introduces these transversions (104). It has been shown that TLS polymerase ζ efficiently bypasses this adduct, but frequently incorporates a mismatched A codon opposite a G codon, thus introducing a G to T transversion (105). *In vivo* homozygous loss of pol ζ in murine cells greatly reduces the survival of AFB1-exposed cells, and their survival could be recovered through exogenous supplementation with human pol ζ (106). Thus, TLS likely plays an important role in cell tolerance to AFB1, and is a probable cause of p53 mutations in situations when NER is unable to act fast enough to remove mutations before replication, and TLS is hence required. Thus, deficiencies in the NER pathway which increase the need for TLS would likely increase AFB1 mutagenesis, and therefore both pathways are important for AFB1 tolerance.

In contrast to the above, multiple lines of evidence also suggest dysregulated NER contributes to liver disease and HCC development independent of AFB1 exposure. By example, some patients with photosensitive Cockayne syndrome develop prominent hepatomegaly and cholestasis (107), and when prescribed metronidazole are at risk of liver failure. Importantly, the two proteins associated with this syndrome, CSA and CSB, are involved in nucleotide excision repair. Moreover, components of the NER pathway associate with HCC. For instance, murine XPA loss causes increased rates of spontaneous liver tumours without AFB1 exposure (101). Similarly, XPC deficiency is associated with increased age-related spontaneous liver mutations in mice (108). In HCC key NER enzymes are overexpressed including XPC, XPA, and ERCC1 (109). In addition, 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3 (PFKFB3), a glycolytic enzyme, translocates into the nucleus of human HCC and increases expression of ERCC1 (110). Connecting to viral hepatitis, HCV-infection increases mutagenesis partly by driving expression of TLS polymerases, thus lowering the overall fidelity of DNA replication (71).

1.3.4.2.3 Variants in nucleotide excision repair may mediate HCC susceptibility

Given the above findings suggesting a relationship between NER and HCC, it follows that polymorphisms may exist in NER proteins that influence human susceptibility to HCC. However, the only well-studied NER polymorphisms are Lys751Gln and Asp312Asn in the XPD protein, and all in patients with high rates of HBV infection (**Table 1.1**). In Chinese populations, the Lys751Gln

polymorphism associated with increased HCC susceptibility, while Asp312Asn was not (83, 86). This association between Lys751Gln and HCC was not observed in a Taiwanese population (84). This discrepancy could be due to the high prevalence of AFB1 exposure in China, or due to population differences in another confounder responsible for HCC-aetiology. To overcome such confounders, meta-analyses have been performed (111, 112). They identified that both variants (Lys751Gln and Asp312Asn) affected HCC susceptibility.

1.3.4.3 Double stranded break repair

Double-stranded breaks (DSBs) are far less common than the previously discussed forms of DNA damage but far more deleterious. The DSB repair (DSBR) pathway is highly disputed in the literature. It is ultimately regulated by three phosphoinositide 3-kinase-related kinases (PIKKs): ataxia telangiectasia mutated (ATM), ataxia telangiectasia and Rad3-related (ATR), and DNA protein kinase catalytic subunit (DNA-PKcs). These proteins phosphorylate H2AX, a histone 2A isoform, and other DNA-repair enzymes to mediate repair (113).

These three PIKKs are recruited to sites of DNA damage by three respective protein complexes. ATM is recruited to DSB ends by the MRN complex, consisting of meiotic recombination protein 11 (MRE11), a 3'-to-5' exonuclease, the DNA binding protein RAD50, and Nijmegen Breakage Syndrome 1 (NBS1), which directly recruits ATM. Similarly, the Ku proteins (Ku70 and Ku80) directly bind DSB ends and recruit DNA-PKcs. In contrast to ATM and DNA-PKcs, ATR does not play a substantial role in the repair of DSB per se, instead it is recruited once single stranded stretches have been created from DSB. Specifically, ATR is recruited to ssDNA by replication protein A (RPA) with help from the 9-1-1 complex (RAD9, RAD1, HUS1) and the ATR-interacting protein (ATRIP) (113, 114).

DSBR occurs via two major pathways, non-homologous end-joining (NHEJ) and homologous recombination (HR). The choice between these pathways is mediated by p53-binding protein 1 (53BP1) and breast cancer susceptibility type 1 (BRCA1), respectively (**Figure 1.10**). NHEJ is the more common pathway and acts as the first line of defence to DSBs. In NHEJ, DNA ends are recognised and bound by the Ku proteins within seconds. These proteins then recruit DNA-PKcs, which in turn phosphorylates H2AX to recruit 53BP1, which suppresses HR by restricting extensive end-resection. Another protein, Artemis, is also phosphorylated by DNA-PKcs, activating its endonuclease activity which performs minor end-processing, finally allowing the DSB ends to be religated by ligase 4 (LIG4) with its binding partners X-ray cross-complementing protein 4 (XRCC4) and XLF (XRCC4-like-factor). This may result in small deletions (115, 116). Alternatively, the presence of ssDNA overhangs at DSBs may also require end-filling by X-family polymerases prior to ligation (117), thus also slightly modifying the sequence around a break

In contrast, the more complex HR requires a homologous template, restricting its activity to the S/G2 phase of the cell cycle when a sister chromatid is available. HR is initiated by the generation of ssDNA

overhangs by exonuclease 1 (EXO1), allowed by BRCA1 activity which blocks 53BP1 suppression of end resection. Immediately, the ssDNA overhang is coated by replication protein A (RPA), which is quickly displaced by RAD51. RAD51, with its binding partner BRCA2, promotes strand invasion of ssDNA tails into a homologous DNA duplex (usually the sister chromatid), creating a primer for new, template-directed DNA synthesis by polymerase δ, κ, or Nu (ν). ssDNA invasion and annealing creates a double Holliday junction (dHJ), which is ultimately resolved through a yet unknown mechanism (115, 118). Finally, an intermediate to these two pathways is the alternative NHEJ (aNHEJ) pathway, which is promoted by Poly(ADP-Ribose) Polymerase 1 (PARP1) and involves limited end-resection to reveal microhomologies. This pathway acts as an alternative to HR (55).

Due to its lack of template-directed repair, introduction of small errors, and potential to indiscriminately religate incorrect DSB ends, NHEJ is classically considered to be the more error-prone pathway. However, over-activity of HR enzymes can lead to a loss of specificity, and an increase in far more deleterious chromosomal rearrangements, mitotic recombination, and loss of heterozygosity (LOH), all features of cancer (43, 119).

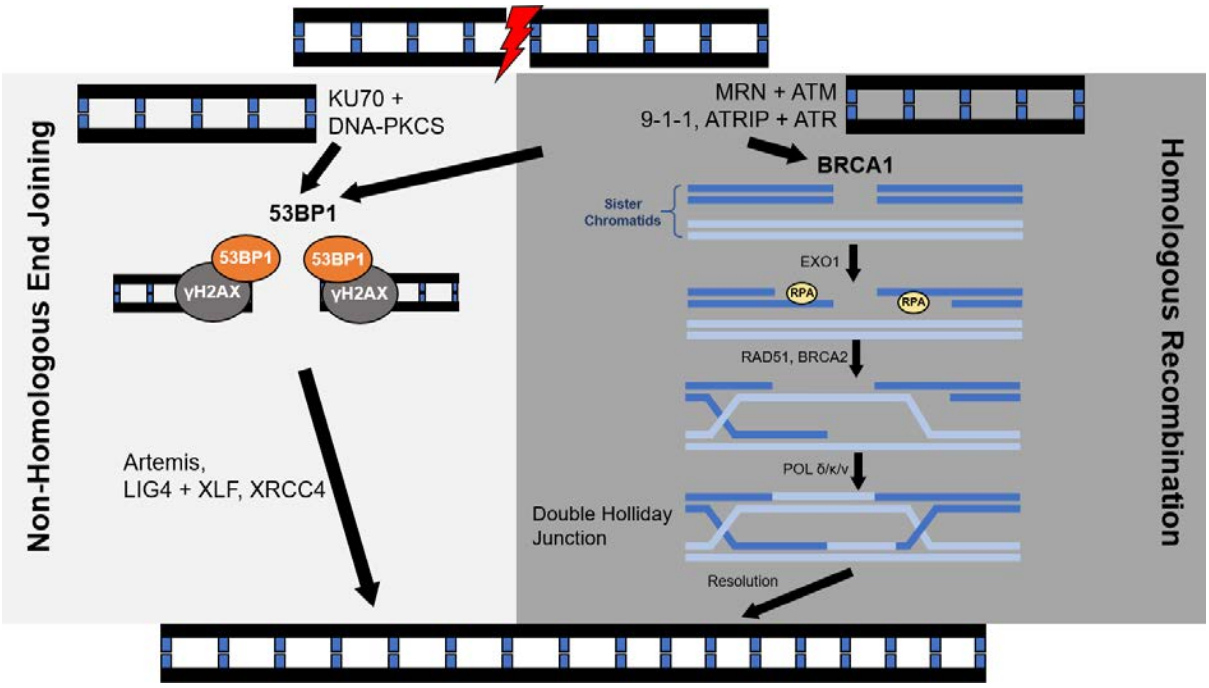

Figure 1.10: Repair of DNA double-stranded breaks (DSBs) occurs via two major pathways: Non-homologous end-joining (NHEJ, Left) and homologous recombination (HR, Right). NHEJ acts as a first line of defence to DSBs and involves relegation of the DNA ends after minimal end-processing. HR is the more complex repair system which is limited to the S/G2 phase of the cell cycle, and provides faithful, template-directed repair using the sister chromatid. See text for detailed description of the specific molecular events.

1.3.4.3.1 Sources of Double-Stranded Breaks in the Liver

DSBs arise in the liver through both exogenous and endogenous sources. On average, spontaneous DSBs occur once every 10^8 bp (120). However, the liver may be prone to higher DSB frequencies than other organs for a number of reasons. Firstly, the liver plays a major role in food and drug metabolism. This leads to the production of toxic metabolites which promote DSBs through oxidative damage (121), that also was discussed above in the context of single-base damage. Nonetheless in terms of DSBs, Enriquez-Cortina, Bello-Monroy *et al.* (66) showed that oxidative stress induced by a high cholesterol diet in mice up-regulated the DSB marker γH2AX. Secondly, the liver is responsible for the production of many large plasma proteins, and this sensitises liver cells to replication fork collisions with transcriptional machinery. This has been suggested as a leading cause of DSBs and genomic instability (120). Indeed, Letouzé, Shinde *et al.* (122) recently showed that mutation rates increase with gene expression, and that the rate of DSB-induced mutations, such as insertions and deletions were enriched in highly expressed, liver-specific genes.

1.3.4.3.2 Double-Stranded Break Repair Activity in HCC

DSBR proteins are commonly up-regulated in HCC tumours, indicating the importance of DSBR to tumour survival. Wang, Li *et al.* (123) found NBS1 was up-regulated in roughly half of all HCCs and cholangiocarcinomas that they investigated. Furthermore, Evert, Frau *et al.* (124) also showed that DNA-PKcs is up-regulated in HCC, and its expression was inversely associated with patient survival. Similarly, Dai, Tang *et al.* (125) showed that EXO1 is commonly up-regulated in HCC (87% of cases), and was likewise associated with poor survival. Additionally, somatic mutations in DSBR proteins are common in HCCs, with a recent study by Lin, Shi *et al.* (126) indicating that ATM, BRCA1, and BRCA2 are all mutated in around 5% of cases. Taken together, this indicates that DSBR alterations likely play important roles in HCC pathogenesis.

Experimentally-induced deficiencies of DSBR have been shown to accelerate HCC progression. For example, heterozygous loss-of-function mutations in NBS1 in mice causes increased susceptibility to a range of cancers including HCC (127). Furthermore, it has been shown that this deficiency can specifically increase hepatic tumorigenesis as compared with other organs (128). Loss-of-function NBS1 mutations have also been observed in HBV-associated HCC (129). Additionally, knockdown of Ku70 increases hepatocarcinogenesis in DEN-treated mice, which display chromosomal aberrations reminiscent of human HCC and increased ATM phosphorylation, likely due to increased DSBs (130). Lastly, an activator of ATR, BIR repeat containing ubiquitin-conjugating enzyme (BRUCE) is commonly deficient in people with hepatitis, cirrhosis and HCC, and results in impaired ATR signalling (131). Interestingly, there is little evidence of deficiencies in proteins specific to HR driving HCC. However, LOH, which may result from overactive HR, is a frequent feature of HCC (132). This

suggests that a loss of NHEJ, resulting in a shift towards HR-mediated DSB repair may be an initiating factor for hepatocarcinogenesis.

A number of studies have also shown that HCV can directly impair DSBR via NBS1. For example, Machida, McNamara *et al.* (133) demonstrated that HCV infection enhances the rates of chromosomal instability in an inflammation-independent manner, due to inhibition of MRN complex formation. This is thought to occur by direct binding of the HCV viral core protein to NBS1. This ultimately leads to impaired activation of ATM and its down-stream mediator, checkpoint kinase 2 (CHK2) (134).

Additionally, HCV and HBV can dysregulate homologous recombination and aNHEJ. For example, the HCV non-structural protein 5A (NS5A) has been shown to interact with RAD51 associated protein 1 (RAD51AP1), which is intimately involved in RAD51-mediated DNA duplex formation for HR. This interaction has a dual function, by inhibiting RAD51AP1 and protecting the viral RNA. While these patients show increased RAD51AP1 expression, the inactivation of the protein results in increased sensitivity to DNA damage due to impaired repair (135). Similarly, an interaction has been demonstrated between the promoter of the HBV core protein and PARP1, which results in both transcriptional activation of the HBV core protein and inhibition of PARP1's DSB repair activity via the aNHEJ pathway (136). This evidence suggests that disruption of DSBR plays an important role in the initiation of HCC.

Finally, HCV can also indirectly prevent DNA damage-induced transcriptional silencing via the induction of ER stress and subsequent down-regulation of Ubiquitin-conjugating enzyme E2 S (UBE2S), whose Lys11-linked polyubiquitination of H2AX chromatin plays a key role in transcriptional silencing (137).

1.3.4.3.3 Variants in Double-Stranded Break Repair May Mediate HCC Susceptibility

A recent study by Lin, Shi *et al.* (126) found germline mutations in a number of DSBR genes, including BRCA1/2, ATM, and RAD50, in 4.2% of patients with primary liver cancer. However, the group only studied a specific subset of genes, and therefore others may have been present. Indeed, genetic studies have found HCC-associated SNPs in a number of DSBR genes (**Table 1.1**). For example, one study found that the presence of a lysine residue in place of a glutamic acid at codon 589 (Glu589Lys) of EXO1 as much as doubled the risk of HCC in a Turkish population (87). Multiple polymorphisms in NBS1 have been investigated as well, with one study indicating an increase in HCC risk associating with Gln185Glu (88). This was somewhat supported by a similar study that found the same SNP to significantly effect HCC risk between HBV-related HCC and HBV infected controls, but not when compared with healthy controls (89), suggesting that this variant could play a role in the mechanism of HBV-related HCC. In addition, slightly protective polymorphisms were found in an intron variant (rs7003908) of the DNA-PKcs gene in males and alcohol drinkers, and the 3' UTR variant rs7963551 in RAD52, in Taiwanese and Chinese populations, respectively (90, 91). Both of these effects were

only small. While it is unclear whether Gly6721Thr affects XRCC7 function, SNP rs7963551 was found to cause up-regulation of RAD52 by preventing miRNA let-7 binding. Thus, this suggests that increased NHEJ activity may be protective for HCC.

1.3.4.4 Mismatched Bases

The most common cause of mutation, and one that arises endogenously, is DNA-mismatch due to incorrect base pairing. Mismatches are repaired via the mismatch repair (MMR) pathway (**Figure 1.11**). MMR is well-studied in prokaryotes, however eukaryotic MMR is poorly understood. Briefly, mismatches are recognised by the heterodimer of eukaryotic Mutator S (MutS) homologs (MSH): either MutSα (MSH2 and MSH6) for small mismatches of 1 to 2 nucleotides, or MutSβ (MSH2 and MSH3) for larger mismatches. MutS then recruits the MutLα heterodimer consisting of DNA mismatch repair protein MLH1 and mismatched repair endonuclease PMS2, which nick the newly-synthesised daughter strand of the DNA duplex. While in prokaryotes this strand discrimination is mediated by hemi-methylated dGATC sites, it is thought that eukaryotic strand-discrimination is mediated by PCNA, though this has not been proven. Nonetheless, the nicked daughter strand is then excised by EXO1, and replaced by DNA polymerase δ (POLδ) and LIG1 (138, 139). Additionally, while not a component of the MMR pathway, the O^6-Methylguanine-DNA-methyltransferase (MGMT) enzyme also helps prevent mismatches by directly repairing O^6-alkylguanine lesions and preventing incorrect mismatching of guanine with thymine (G:T) (140).

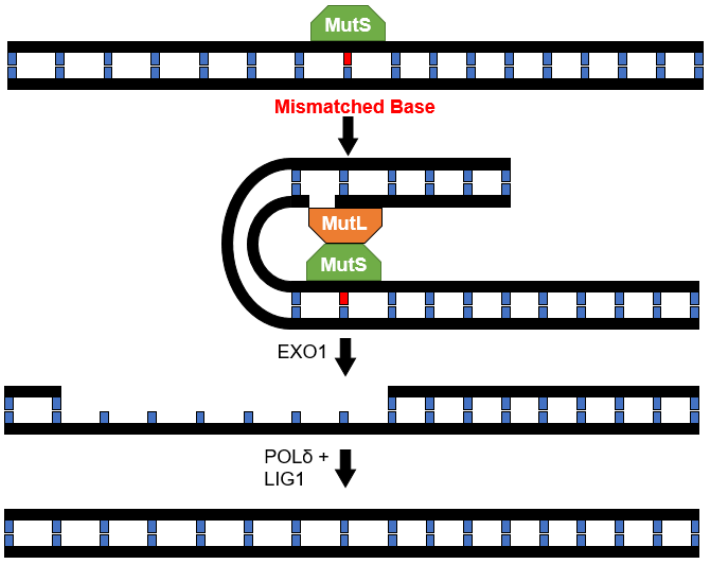

Figure 1.11: Repair of DNA mismatches via the mismatch repair pathway (MMR). The molecular mechanisms of this pathway are poorly understood in eukaryotes, however the process involves recognition of the mismatch by eukaryotic MutS homologs, excision of the mismatched region by MutL homologs and exonucleases, and repair by polymerases and DNA ligases. See text for molecular description.

1.3.4.4.1 Mismatch Repair in the Liver

Given that mismatches arise during replication, MMR is of particular importance in the regenerative liver during chronic liver diseases. In the absence of disease, the liver is composed of quiescent hepatocytes. However, unlike other organs, the liver is able to regenerate itself after damage in a process called compensatory hyperplasia, where the remaining hepatocytes undergo mitosis (141). As discussed previously, chronic inflammation through the production of factors such as TNFα and IL-6 by Kupffer cells, is a common feature of liver diseases regardless of their aetiology. While this inflammation induces cell death, it also promotes proliferation to regenerate lost cells (74). This leads to a dramatic increase in cell turnover rate, which in turn increases the likelihood of DNA-mismatches.

1.3.4.4.2 Mismatch Repair and Microsatellite Instability in HCC

The data remains unclear whether MMR plays a role in HCC. It is known that MMR deficiencies result in microsatellite instability (MSI). Microsatellites are small DNA motifs consisting of 1-6 nucleotides that can be repeated many times (142). MSI is seen much less frequently in HCC when compared with other forms of genomic instability like LOH (132). Takagi, Esumi *et al.* (143) reported MSI in 29% of HCCs in patients who had multiple primary tumours. However, when patients presented with only HCC, MSI was only found in 6% of cases. This suggests that MSI is not specific to HCC. In line with these findings, three studies from America (144, 145) and England (146) found no involvement of MSI or reduced MMR protein expression in non-HCV-associated HCC. Furthermore, patients with lynch syndrome caused by mutations in the major mismatch repair enzymes (MSH2, MSH6, MLH1, PMS2) do not typically develop HCCs; only one publication exists of a HCC in a patient with lynch syndrome (147).

In contrast to the above, some studies specifically suggest HCV-related HCCs are linked with MMR deficiency. Two Egyptian studies identified frequent (70-90%) MMR deficiencies, most often in MSH2 and MLH1, in HCV-related HCCs, whilst HCV-infected individuals without HCC had normal MMR expression in their liver tissue (148, 149). Additionally, Wani, Notohara *et al.* (150) studied HCV-associated HCC in a Japanese cohort and found that decreased MSH2 and MLH1 staining correlated with poor to moderate differentiated tumours, but not with well-differentiated tumours, suggesting that loss of MMR plays a role in early HCV-associated HCC. This is consistent with a bioinformatic, protein-protein interaction network study which also suggested MMR involvement in early HCC (151).

Thus, these findings suggest MMR deficiencies may be associated with HCV-related HCC, but not with non-viral HCC. This could be due to such deficiencies facilitating viral infection of hepatocytes, or could be due to viral proteins directly down-regulating MMR activity as discussed in other DNA-repair pathways. The role of MMR in other forms of HCC is less clear. However, almost all of the studies described have assessed MMR expression only via immunohistochemistry, rather than by more

sensitive techniques such as RNA sequencing or quantitative PCR. Thus, further studies are still required to evaluate the role of MRR in HCC.

1.3.4.4.3 Mismatch Repair Variants in HCC Susceptibility

Given the potential role of MMR in HCC, germline polymorphisms in MMR proteins could predispose individuals to HCC, particularly in the case of HCV-related HCC. However, only three articles have explored this, one in MSH2 (92) (Chinese) and two in MLH1 (84, 93) (Chinese and Taiwanese), all of which focussed on HBV-related HCC (**Table 1.1**). Zhu, Wang *et al.* (92) studied 8 polymorphisms in the MSH2 gene and found only one that significantly affected HCC risk, an intron variant with an odds ratio of approximately 1.8. Similarly, Zhu, Liu *et al.* (93) investigated 4 polymorphisms in MLH1 and reported a single 5'-UTR variant (rs1800734) which significantly increased HCC risk. This was supported by Chen, Yang *et al.* (84) who reported rs1800734 as the only polymorphism which significantly affected HCC risk in a panel of polymorphisms which included inflammatory cytokines and other DNA repair proteins. Finally, the previously discussed study by Bayram, Akkız *et al.* (87) found significantly increased HCC risk in individuals with a Glu589Lys polymorphism in EXO1, which is known to interact with a number of other MMR proteins to potentially stabilise the MMR complex (152).

No studies to date have investigated an association of MMR with HBV-related HCC outside of these genetic variants. These results, though limited, further suggest a potential role of MMR disruption in viral HCC, here specifically with HBV infection, which requires the support of experimental studies to investigate further.

1.3.4.5 p53 in HCC

The p53 tumour suppressor protein is a master regulator of numerous cellular processes, including DNA damage repair and the cell's response to DNA damage. p53 is the most frequently mutated protein in cancer cells. Indeed, p53 is mutated in 50% of aflatoxin-induced HCC, and 28-42% of non-aflatoxin induced HCC (153). A previous review (154) described the roles of p53 in DNA-repair. Briefly, it has been discovered that p53 is capable of regulating all major DNA repair pathways, both through transcriptional regulation of DNA-repair enzymes, and through transcription-independent means. While p53 appears to have some link to every repair process, its effects are most obvious in HR, and least obvious in MMR and NHEJ. Thus, it is likely that many of the roles of DNA repair in HCC discussed in this review may be secondary effects of altered p53 activity.

1.3.4.6 The DNA Damage Mechanism of Liver Cancer

Given the evidence presented, DNA damage and subsequent repair clearly plays an important role in the liver, and is likely to be critical in the initiation and progression of liver cancer. There is strong evidence that alterations in any of the five major repair pathways can affect liver cancer progression.

Thus, based on the presented evidence a general mechanism of DNA damage-related liver cancer can be described (**Figure 1.12**).

In this mechanism, DNA damage-related HCC occurs through two key steps. Initially, a DNA repair pathway is impaired by a diverse range of factors depending on the aetiology of disease. This can include: direct interactions with viral proteins causing inhibition (79, 133, 135-137); direct interactions with inflammatory compounds (64); somatic mutations (126, 155); epigenetic down-regulation (156, 157); and congenital, germline mutations. In many cases, the HCC-initiating potential of these impairments has been demonstrated through experimental studies.

Deficiencies in any repair pathway increase the likelihood of persistent mutations, which can cause inactivation of tumour suppressors or activation of oncogenes. However, due to the pathogenesis of different aetiologies of HCC, certain pathways become more relevant with different risk factors. BER deficiencies cause genomic instability in inflammatory environments, as does DSBR, due to the effects of ROS on DNA. In terms of DSBR, cancer initiation appears to be driven primarily by deficiencies in NHEJ rather than HR. NER deficiencies are most relevant to DNA-intercalating compounds like AFB1. Finally, MMR deficiencies appear to be the least relevant to non-viral HCC, with their effects primarily evident in HCV-induced HCC. The reason for this remains to be elucidated.

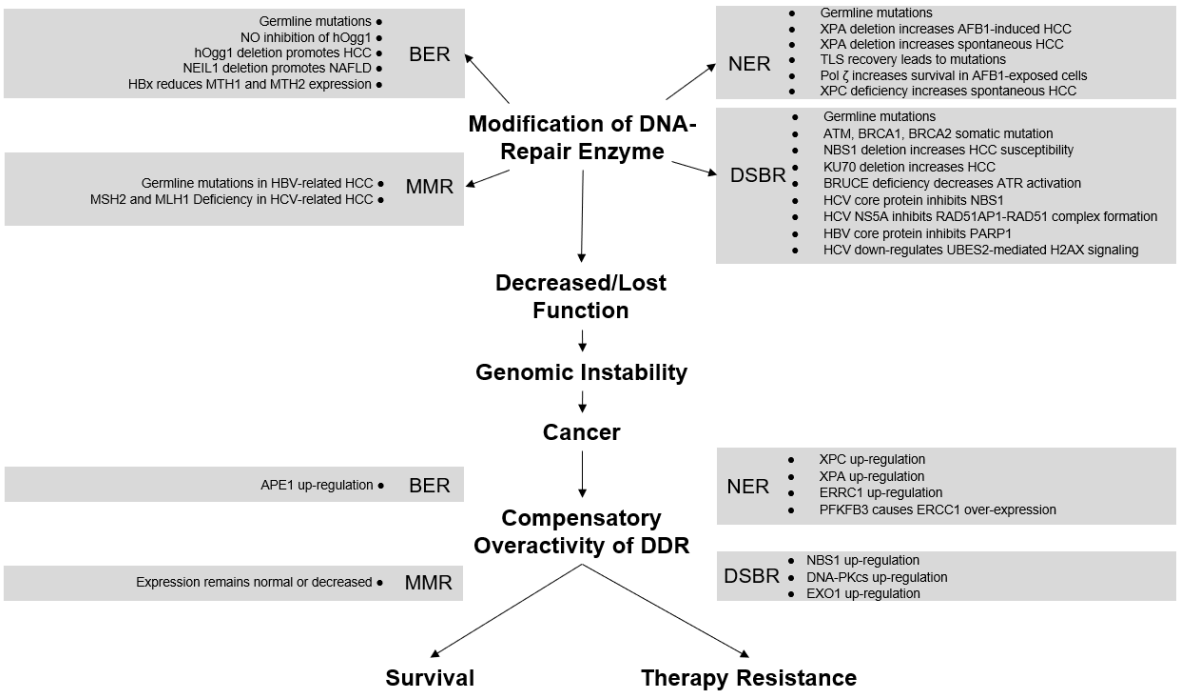

Figure 1.12: The DNA-damage mechanisms of liver cancer. The initiation of liver cancer can be driven by numerous events including deficiencies in DNA repair pathways that result in mutagenesis. In turn, these deficiencies drive the over-expression of recovery DNA-repair pathways, which drive cancer survival and therapy resistance.

Following these initiating deficiencies, cancer cells require alternative mechanisms to retain the stability required for survival in the harsh tumour environment, and to inadvertently resist DNA-damaging chemotherapeutic agents. This is supported by the fact that DNA-repair enzymes are often found to be over-expressed in HCC. DNA repair pathways are often semi-redundant and capable of recovering another pathway's deficiency to an extent (56). Over-expression of enzymes is evident in almost every pathway except for MMR. A reason for this could be that the MMR pathway is less effective in compensating for other pathways, however this needs to be investigated further. Nonetheless, over-expression of DNA-repair enzymes is likely to be responsible for the notorious chemo- and radio-resistance of HCC (158).

1.3.4.7 Potential for HCC Treatment

HCC is highly resistant to most classical chemotherapies, including those which work by inducing DNA damage, such as cisplatin. A recent review by Marin, Macias *et al.* (159) thoroughly discussed the limited current state of knowledge regarding the mechanisms behind this resistance. Because of this, the approved first-line drug therapies for HCC include Sorafenib, followed by Regorafenib as an approved second-line therapy, and lenvatinib for treatment of unresectable hepatocellular carcinoma, all of which are broad-acting multi-kinase inhibitors that mainly target cell proliferation. Moreover, these drugs are non-curative and extend survival by just 10-14 months and can have adverse effects (160).

The proposed DNA damage mechanism of liver cancer presents a tantalising opportunity for targeted therapy towards cancer cells, due to an effect known as 'synthetic lethality'. When cancers arise due to deficiencies in DNA repair, and these pathways are recovered by other redundant repair pathways, they become highly sensitive to therapeutics targeting those recovery repair pathways. This means that pharmacological inhibitors are synthetically lethal to these cancer cells, while they have little effect on functionally normal cells (55). This therapeutic strategy has been primarily used in breast cancer caused by BRCA mutations using PARP inhibitors. Furthermore, evidence suggests that while most cancer types tend to have co-occurrence of mutations in DNA-repair enzymes, liver cancer is unique in that DNA-repair mutations are generally mutually exclusive (161). This suggests that loss of multiple DNA repair pathways is particularly detrimental to liver tumours, meaning that they would be hypersensitive to synthetically-lethal drugs.

At the time of writing, the overwhelming majority of current clinical drug trials listed in clinicaltrials.gov relate to broad kinase inhibitors, growth factor inhibitors, anti-angiogenic drugs, or immunotherapies. Of the 2086 trials listed, table 2 describes the six drug trials with relevance to DNA repair.

Two clinical trials have investigated Veliparib (ABT-888) in combination with Temozolomide for HCC. Veliparib is a PARP inhibitor, which is hoped to induce synthetic lethality in combination with

Temozolomide, an alkylating agent which induces DNA damage requiring PARP-dependent repair. The first of these trials (NCT00526617) concluded that there was no adverse pharmacokinetic interaction between the drugs (162). The second trial (NCT01205828) showed that the drugs were mostly well-tolerated in patients with advanced HCC, but was terminated due to a lack of efficacy (163). This was likely because the treatment did not appear to be at all specifically targeted to patients with the relevant DNA repair deficiencies. Additionally, one currently active clinically trial (NCT01356628) is investigating Palbociclib (PD-0332991) for HCC. This drug, while primarily a CDK 4/6 inhibitor, has been found to also inhibit ATM and enhance radiosensitivity (164), which suggests potential application as a synthetically lethal drug.

The remaining trials listed in **Table 1.2** are included because they involve novel use of DNA-damaging agents for HCC treatment. One of these trials (NCT01752933) is for SGI-110, a novel DNA methylation inhibitor which has been shown to increase cisplatin-induced DNA damage in ovarian cancer cells (165). The final two trials (NCT01775501, NCT02042443) are for Fluorouracil, or Capecitabine which is metabolised into Fluorouracil. Fluorouracil is an uracil analogue which inhibits nucleotide production to thus prevent DNA replication and repair. The drug has been shown to induce DNA damage, and trigger the ATM/checkpoint kinase 2 (CHEK2)/p53 pathway in colorectal cells (166). However, no results have been published on any of these three drug trials at the time of writing.

The limitation of such therapeutic targeting is that it requires a high degree of personalisation, but individual patient analysis to identify deficient pathways and use of three-dimensional tumor organoids, may lead to the development of highly effective therapies. Furthermore, in using DNA repair inhibitors to treat cancer, researchers and clinicians should be cautious of the risk of further triggering tumorigenesis and progression.

Table 1.2: Clinical trials of DNA repair- and DNA damage-targeted therapies. Information was retrieved from clinicaltrials.org.

NCT Number	Title	Drug of Interest	Enrolment	Conditions	Phase	Funded By	Country	Status	Year	Related Publications
NCT00526617	A Phase I Study of ABT-888 in Combination With Temozolomide in Cancer Patients	ABT-888 (Veliparib) and Temozolomide	41	Non-hematologic Malignancies	Phase 1	Industry	US	Completed	2007-2017	(162)
NCT01205828	ABT-888 and Temozolomide for Liver Cancer	ABT-888 (Veliparib) and Temozolomide	16	Hepatocellular Carcinoma	Phase 2	Industry	US	Terminated	2010-2016	(163)
NCT01356628	A Clinical Research Study to Determine Whether PD 0332991 May Be Effective in Treating Patients With Liver Cancer	PD-0332991 (Palbociclib)	23	Hepatocellular Carcinoma	Phase 2	Industry	US	Active	2011-Current	None
NCT01752933	SGI-110 in the Treatment of Advanced Hepatocellular Carcinoma (HCC)	SGI-110	52	Hepatocellular Carcinoma	Phase 2	Industry	US	Completed	2012-2019	None
NCT01775501	Sorafenib + mFOLFOX for Hepatocellular Carcinoma	FOLFOX Drug Combination (Leucovorin, Fluorouracil, Oxaliplatin) and Sorafenib	40	Hepatocellular Carcinoma	Phase 2	Other	US	Completed	2013-2020	None
NCT02042443	Trametinib or Combination Chemotherapy in Treating Patients With Refractory or Advanced Biliary or Gallbladder Cancer or That Cannot Be Removed by Surgery	Capecitabine	53	Many Cancers	Phase 2	NIH	US	Completed	2014-2017	None

1.3.4.8 Conclusion

This review aimed to investigate the role of DNA repair pathways in the development of liver cancer. As the liver is exposed to unusually high levels of DNA-damaging stimuli, DNA repair is crucial to preventing cancer. Indeed, deficiencies in BER and DSBR appear to drive much carcinogenesis. As expected, NER deficiencies appear to primarily drive AFB1-related cancer, however there is evidence to support NER importance in spontaneous liver carcinogenesis as well. MMR deficiencies, surprisingly, appear to have the least impact in carcinogenesis, with the only strong evidence of their importance demonstrated in HCV-related HCC. Many genetic studies have been performed in order to identify genetic variants in DNA-repair genes which may drive HCC susceptibility. However, these studies produced highly variable results, and generally only show small effects in HCC risk. An explanation for this could be that polymorphisms in DDR genes in general can increase susceptibility to HCC, but an individual analysis of any one in isolation lowers the predictive ability of the test. Additionally, different populations may present different susceptibility variants associated with different risk factors. For this reason, future approaches to such genetic studies could focus on a pathway-wide view of repair enzymes, rather than specific genes in isolation.

1.3.5 Statement of Intent

HORMAD2 is a poorly characterised protein with preliminary data suggesting a completely novel role in HCC. HORMAD2 is known to function in meiotic pathways which closely parallel the DNA-damage response to double-stranded breaks in somatic cells, and interacts either directly or indirectly with a number of DSB-related proteins. Thus, it is hypothesised that HORMAD2 may influence HCC progression through manipulation of DNA-repair processes. The aims of this component of the project were therefore:

1. **Characterise HORMAD2 *in silico* and investigate its significance to human HCC through bioinformatic analysis of human HCC datasets and other data available online. This will include the use of gene expression correlation and ontology analysis, and the analysis of HORMAD2 expression and its relationship with patient survival and clinical presentation.**
2. **Use the findings of Aim 1 to guide an *in vitro* characterisation of HORMAD2 in HCC. This will be achieved primarily through the use of knock-down and overexpression constructs in HCC cell lines, as well as biochemical assays utilising recombinant HORMAD2 protein.**

1.4 Oral-Pharyngeal Cancer

Cancers of the lip, oral cavity and oropharynx (Oral-Pharyngeal Cancer; OPC) are among the 10 most common cancers in the world (167), and within Queensland, Australia the incidence of these cancers is increasing annually by 10% (168). The main risk factors for OPC include tobacco use and infection with the Human Papilloma Virus (HPV). Over the past few decades, very little progress has been made in improving the overall survival of OPC patients. With decreasing rates of tobacco use globally, research efforts have shifted away from tobacco use, and instead towards HPV infection (167) and de-intensifying chemoradiation therapy in HPV infected patients given it better prognosis. However, smoking rates in regional and remote areas of Australia remain almost double that of urban centres and major cities (21% vs 13%) (169). Furthermore, non-HPV-related OPC is associated with far worse survival, and continual smoking during radiotherapy further decreases survival and promotes metastasis clinically (170-172). Thus, for the benefit of regional and remote Australian public health, it is of great importance to continue to elucidate the mechanisms of increased severity in smoking-related OPC.

1.4.1 Treatment

Surgical resection and radiation therapy, with or without chemotherapy, remain the frontline treatment options for OPC (173). However, radiation therapy has been shown to induce dedifferentiation of HPV-negative OPC cells into cancer cells which express stem cell markers such as CD44 and ALDH (172, 174). Additionally, radiation treatment for OPC is known to be a risk factor for development of other primary cancers in the oral cavity, pharynx, oesophagus, and lung (175). For this reason, it has been understood for decades that sub-curative treatment of cancers promotes more aggressive and invasive phenotypes (176, 177), hence the need for follow-through treatment.

1.4.2 Metastasis

Metastasis is the primary cause of death in most malignant cancers. It is a multi-step process which requires a number of phenotypic changes to occur (178); cancer cells detach from the primary tumour site, invade through a basement membrane, circulate through lympthatics or circulatory system – evading the immune system, attach to endothelial cells, invade through a second basement membrane and seed the growth of secondary tumours at distant sites. Together, these stages make up the ‘metastatic cascade’.

Local invasion is the key initiating step of metastasis. This is achieved through deadhesion of the tumour cells from their surrounding tissue, mediated by changes in the expression of cellular adhesion molecules, enhanced motility of the cell itself, as well as the secretion of proteases which facilitate the degradation of the extracellular matrix and basement membrane (178). The key adhesion molecules involved in local tumour invasion include integrins which attach cells to the extracellular matrix (ECM), and cadherins which maintain cell-cell adhesion. Integrins involved in the maintenance of stable adhesion are often lost in tumours, however those involved in cell migration may be up-regulated (179).

The acquisition of enhanced cell motility is extremely complex, however at its core it is mediated by the cytoskeletal reorganisation of Rac and Rho family GTPases, which restructure the actin cytoskeleton to allow the formation of lamellipodia (180). Finally, invasion also relies on the degradation of surrounding ECM and basement membrane. This is mediated by the secretion of proteases, most notably in the form of metalloproteases (MMPs) (181), and cathepsin family proteases (182).

Intravasation describes the process of entrance into the blood circulation. This relies again on the previously described enhancement of cellular motility, but also on the accessibility of blood vessels which is mediated through angiogenesis, the process of creating a primitive disorganised circulatory system. Enhanced angiogenesis in tumours is usually triggered by the expression and release of angiogenetic growth factors like the vascular endothelial growth factors (VEGF), which are released by hypoxic cells such as those found within solid tumours. These factors signal the development of microvasculature around the tumour (181).

Extravasation involves the tumour cell's liberation from the blood stream and deposition into the tissue of a distant site. However, this deposition is inconsequential without the capacity for a stem cell to seed the formation of an entire tumour. This property is described as stemness, and is analogous to the properties of normal stem cells, only they exist as their malignant equivalents (183).

1.4.3 Cancer Stem Cells

Cancer stem cells (CSCs) are a key component of the metastatic cascade. CSCs describe a population of cells within a tumour that are capable of self-renewal and seeding of new cancers. Importantly, CSCs are more resistant to chemotherapeutics and radiation (174, 184, 185). Furthermore, radiation treatment has been shown to increase the proportion of CSCs within a tumour (186).

Numerous cell-surface markers, intracellular enzymes and transcription factors have been identified as markers of these CSC phenotypes. For example, CD44 is a transmembrane glycoprotein that acts a ligand for hyaluronan and stimulates pathways involved in cell proliferation and cellular motility (187). Additionally, aldehyde dehydrogenases (ALDH) are cytosolic enzymes involved in the oxidation of aldehydes. As stem-cell regulators, ALDH enzymes enhance a cells capacity for DNA repair and reduce ROS, improving survival and resistance to therapies, and they also indirectly regulate gene expression through retinoic acid-dependent pathways (188).

1.4.4 Continued smoking during radiation treatment

Clinically, patients undergoing radiotherapy for OPC who continue to smoke have been shown to have dramatically decreased survival and locoregional control of their cancers when compared with patients who quit smoking prior to therapy. It was suggested that the causes of these worsened outcomes were related to smoking's effects on chronic hypoxia which decrease the efficacy of radiation treatment (171). However, *in vitro* exposure of OPC cancer cells to cigarette smoke has been shown to induce a

CSC-like phenotype (189). Thus, it is possible that through unknown mechanisms, cigarette smoke is able to induce changes in cell invasiveness and stemness on a molecular level which does not require an *in vivo* model.

1.4.5 Statement of Intent

Patients who continue to smoke while undergoing radiotherapy treatment for OPC suffer worse survival and increased likelihood of metastasis. While it has not been established, evidence suggests that these effects may be occurring at a molecular level, rather than them being reliant on the effects of hypoxia in an *in vivo* setting, as previously suggested. The aims of this component of the project were therefore:

1. **Analyse the expression of genes in OPC cell lines after treatment with radiation, cigarette smoke, or combination treatment of both. Pathway enrichment analysis will be used to identify the genetic pathways most affected by each of these treatments.**
2. **Identify genes responsible for observed changes in tumour invasiveness. More specifically, curated genetic pathways of tumour invasiveness will be investigated to identify hit genes.**

Chapter 2: Materials and methods

2.1 Bioinformatics

2.1.1 Software

Table 2.1: Bioinformatics software.

Name	Version	Reference
DESeq2	1.30.0	(190)
tximport	1.18.0	(191)
Tidyverse	1.3.0	(192)
Survival	1.18.0	(193)
multiMiR	1.12.0	(194)
FastQC	0.11.7	(195)
MultiQC	1.9	(196)
RSEM	1.3.1	(197)
ggVennDiagram	0.5.0	(198)

2.1.2 HORMAD2

2.1.2.1 TCGA-LIHC

2.1.2.1.1 Publicly Available Datasets

The data used for this analysis was generated by the Cancer Genome Atlas (TCGA) research network https://www.cancer.gov/tcga. All data analysed came from the liver hepatocellular carcinoma (LIHC) cohort. As all TCGA data is publicly available without restricted use, there are numerous sources from which the data can be accessed. This includes the raw data direct from the National Cancer Institute (NIH) Genomic Data Commons (GDC) Data Portal, as well as partially pre-processed data available from various institutions such as the Broad Institute and the University of California Santa Cruz (UCSC).

Pre-processed TCGA-LIHC RNASeq, mutation and patient clinical data was retrieved from the Broad Genome Data Analysis Centre (GDAC) Firehose via the Firebrowse data portal (http://firebrowse.org/?cohort=LIHC). This included data for 371 patients, 50 of which had paired normal liver tissue RNASeq data.

Pre-processed TCGA-LIHC methylation data was retrieved from the University of California Santa Cruz (UCSC) Xena platform.

Table 2.2: Data sources for human HCC bioinformatics.

Data Type	Source
TCGA-LIHC RNASeq Data	http://gdac.broadinstitute.org/runs/stddata__2016_01_28/data/LIHC/20160128/gdac.broadinstitute.org_LIHC.Merge_rnaseqv2__illuminahiseq_rnaseqv2__unc_edu__Level_3__RSEM_genes_normalized__data.Level_3.2016012800.0.0.tar.gz
TCGA-LIHC Clinical Data	http://gdac.broadinstitute.org/runs/stddata__2016_01_28/data/LIHC/20160128/gdac.broadinstitute.org_LIHC.Merge_Clinical.Level_1.2016012800.0.0.tar.gz
TCGA-LIHC Mutation Data	http://gdac.broadinstitute.org/runs/stddata__2016_01_28/data/LIHC/20160128/gdac.broadinstitute.org_LIHC.Mutation_Packager_Oncotated_Raw_Calls.Level_3.2016012800.0.0.tar.gz
TCGA-LIHC Methylation Data	https://tcga.xenahubs.net/download/TCGA.LIHC.sampleMap/HumanMethylation450.gz https://tcga.xenahubs.net/download/probeMap/illuminaMethyl450_hg19_GPL16304_TCGAlegacy
TCGA-LIHC miRNA Data	http://gdac.broadinstitute.org/runs/stddata__2016_01_28/data/LIHC/20160128/gdac.broadinstitute.org_LIHC.Merge_mirnaseq__illuminahiseq_mirnaseq__bcgsc_ca__Level_3__miR_gene_expression__data.Level_3.2016012800.0.0.tar.gz

2.1.2.1.2 Differential Expression and Gene Count Normalisation

The analysis of differential RNA expression in the TCGA-LIHC cohort was performed based on negative binomial distribution using the R/Bioconductor package 'DESeq2' version 1.30.0. Briefly, independent filtering is performed to select a set of genes for multiple test correction which maximizes the number of adjusted p-values less than the alpha value. Outliers are then removed based on Cook's distance. P values are calculated using a Wald test, and adjusted using the Benjamini & Hochberg correction. Significant differential expression was determined as a DESeq2 adjusted p value < 0.05. Because paired tumour-normal samples were available for only a fraction of the cohort, both a paired and unpaired differential expression analysis was performed. Due to improved statistical power, the paired analysis was used for differential expression analysis.

Additionally, DESeq2 was utilized to normalize read counts based on sequencing depth and RNA-composition using a median ratios method. These normalized reads were used for further analyses such as expression correlation analysis, survival analysis, and expression correlation with patient clinical data.

2.1.2.1.3 Cohort Separation Based on HORMAD2-Expression

In order to investigate the effects of HORMAD2 expression on patient outcome, the TCGA-LIHC cohort was separated into three groups based on whether HORMAD2 expression in each tumour sample was lower than, higher than, or within 1 standard deviation of the mean of HORMAD2 expression in normal liver tissue.

2.1.2.1.4 Survival Analysis

For each patient, vital status was acquired from the clinical dataset. Survival time for deceased patients was taken as days until death. Survival time for non-deceased patients was censored using time until their last follow up. Only patients with both HORMAD2 expression data and survival data were kept

for analysis (n = 370). Survival analysis was performed using the R package 'Survival', with patients separated into High, Medium and Low HORMAD2 expression cohorts.

2.1.2.1.5 Mutation Analysis

As a measure of genomic stability, the frequency of mutations was retrieved from the mutation data and grouped by mutation type: deletions, insertions and single nucleotide polymorphisms (SNPs).

Additionally, mutation data for each gene per patient was retrieved and filtered to remove all silent mutations. The Fisher's Exact test ($p = 0.05$) was used to determine significant differences in mutation frequency between the high vs normal, and low vs normal HORMAD2 expression groups. Genes with significantly enhanced mutation frequencies associated with high or low HORMAD2 expression ($p < 0.05$) were retrieved. These lists were analysed for pathway enrichment of gene ontology (GO) Biological Processes using the Gene Ontology enRIchment anaLysis and visuaLizAtion tool (GORILLA) (199), where the list of significantly mutated genes was submitted as an unranked list, and a list of all genes in the dataset was used as background.

2.1.2.1.6 Co-Expression Analysis

Co-expression analysis was performed using Spearman correlation. In R, Spearman correlation analysis was performed for the DESeq2-normalised RNA expression of HORMAD2 versus every other gene in the dataset. These correlations were then ranked by the strength of their correlation based on the Spearman Rho value. Ranked lists were submitted to the GORILLA database to identify pathway enrichment of GO biological processes. The outputted pathways of the GORILLA search were then submitted to Reduce and Visualise Gene Ontology (REVIGO) using the "small (0.5)" setting for allowed similarity to remove redundant pathway results.

2.1.2.1.7 Methylation Analysis

Partially processed methylation data for the TCGA-LIHC cohort was retrieved from the UCSC Xena platform (https://xenabrowser.net/) including methylation levels at each CpG site per patient in the form of beta value, and CpG mapping information for the Hg19 genome. The beta value is an estimate of methylation based on the ratio of intensities between completely methylated and unmethylated alleles, given as a value between 0 and 1 (0 being unmethylated and 1 being fully methylated). As the methylation values were non-normally distributed, methylation levels were compared with tumour and normal samples, and between HORMAD2-expression groups using the Wilcoxon test and Kruskal-Wallis test with Dunn's multiple comparison test respectively. Correlation was performed using Spearman correlation.

2.1.2.1.8 miRNA Analysis

miRNA expression data for TCGA-LIHC was retrieved from the GDAC firehose pipeline. miRNA expression data was normalised using the R/Bioconductor package 'DESeq2'. Spearman correlation was used to correlate the normalised HORMAD2 RNA expression of each patient with each miRNA.

To further narrow down results, the R/Bioconductor package 'multiMiR' was used. This package automatically searches 15 databases for miRNAs which target the HORMAD2 gene. This includes databases for both validated miRNA targets and predicted targets. The correlation results were then filtered to only include those with evidence of HORMAD2 targeting.

2.1.2.2 Homology-Based Structural Modelling

Several web-based servers for homology-based structural modelling were utilized to generate predicted 3-dimensional structures of the HORMAD2 protein. These included I-TASSER, Phyre2, ROBETTA, and SWISS-MODEL. Models were predicted using the HORMAD2 protein sequence from UniProt (Q8N7B1).

The top scoring model from each homology-based structural modeler was then analysed using the Structure Analysis and Verification Server (SAVES).

2.1.2.3 Transcription-Factor Binding-Site Prediction

The HORMAD2 promoter sequence was retrieved by extracting the 2kb upstream sequence from the beginning of exon 1 of HORMAD2 (GRCh38 chr22:30081011). This sequence was then given as input to the transcription factor binding site prediction tool PROMO V3 (200, 201), limited to only human transcription factors and to a dissimilarity rate threshold of 5%.

2.1.2.4 Data-Mining for HORMAD2 Molecular Interactions

Three different databases were searched for HORMAD2 molecular interactions: Interactome3D (202), Targetmine (203), The BioGrid (204). These databases document the results of high-throughput protein-protein interaction studies.

2.1.2.5 Protein Features

The HORMAD2 amino acid sequence was used at input into a number of online tools which analyse amino acid sequences and identify and predict various features. The tools utilised for this investigation were: NLSdb (205), NLS mapper (206), ELM (207), ProTStab (208), and PredictProtein (209). All searches were performed using default parameters.

2.1.3 Oral Pharyngeal Cancer

2.1.3.1 High-Throughput Sequencing

RNA samples were prepared and sent away for high-throughput Illumina sequencing by the Australian Genome Research Facility (AGRF). Image analysis was performed by the NovaSeq Control Software

v1.7.0 and Real Time Analysis v3.4.4. The Illumina bcl2fastq 2.20.0.422 pipeline was used to generate the sequence data, which was deemed to meet AGRF quality standards. In total, 32 to 57 million paired-end reads were acquired per sample for the 12 samples. All raw files were quality checked using FastQC v0.11.7. and MultiQC v1.9. Sequence quality, and adapter content were all above relevant quality thresholds.

2.1.3.2 Read Alignment

To quantify reads, read alignment was performed using Bowtie2 within RSEM v1.3.1 to the Human Transcriptome GRCh38p.13, retrieved from NCBI. Expression calculation was performed for paired-end reads using default parameters. Quality checking of alignment involved analysis of output flagstat files and FastQC. All samples passed QC and were 94-96% mapped successfully.

2.1.3.3 Normalisation and Differential Expression

Normalisation and differential expression analysis were performed using the R/Bioconductor package 'DESeq2' version 1.30.0. Differential expression was performed as pair-wise comparisons between the 4 treatment groups. Significance was determined using the Wald test with a $\log_2$-fold change threshold of $\log_2(1.25)$ to identify differentially expressed genes with at least a 25% change in gene expression, and identified as $p < 0.05$. Additionally, DESeq2 was utilized to normalize read counts based on sequencing depth and RNA-composition using a median ratios method.

2.1.3.4 Differential Expression Pathway Enrichment Analysis

For each pairwise differential expression analysis, pathway enrichment was performed on the entire list of genes outputted from DESeq2 ranked by the Wald "stat" value and submitted to the GORILLA tool (199) to identify pathway enrichment of GO biological processes. The outputted pathways of the GORILLA search were then submitted to REVIGO (210) using the "small (0.5)" setting for allowed similarity to remove redundant pathway results.

2.1.3.5 Heatmaps

Heatmaps were created by hand-selecting genes based on curated genetic pathways as discussed in the text. For gene expression values, DESeq2's variance stabilising transformation was used which log-normalises the data and stabilises the changes in variance associated with changes in the mean. Values were then centred by subtracting the mean gene-by-gene from each value. Therefore, the final values are a normalised log2 fold-change in gene expression.

2.2 Mammalian Cell Culture

All human cell lines (Huh7, Hep3B, PLC/PRF/5, HEK293T) were cultured in high glucose DMEM cell culture media (D6429, Sigma Aldrich) supplemented with 10% heat-inactivated FBS (35-076-CV, Corning). Cell were maintained at 37°C with 5% CO_2. For MTT assays comparing the effects of cell culture media sugar constitution, the previously stated high glucose DMEM + FBS was used as a

control, while glucose-free DMEM (11966-025, Gibco) was supplemented with glucose or fructose for the experimental conditions. For MTT and comet assays with H_2O_2, FBS was removed from the media and cells were cultured in only high glucose DMEM with H_2O_2 for the duration of H_2O_2 exposure.

2.2.1 Storing Cells

For storage of cell lines following cell culture, cell culture media was aspirated and cells were washed twice with PBS. Cells were lifted from cell culture plates enzymatically by incubation with trypsin for 5 minutes at 37°C, then pelleted by centrifugation at 220 x g for 2 minutes. Cells were resuspended in FBS (35-076-CV, Corning) with 10% DMSO to a concentration of approximately 1-2 million cells/mL. Cells were distributed into 1mL aliquots in cryotubes and brought to -80°C slowly by wrapping tube racks in towels.

2.2.2 Antibiotics

The shRNA knockdown constructs and PCW57.1 inducible-overexpression constructs (discussed later) provided puromycin resistance to the Hep3B and Huh7 cell lines that they were integrated into. Kill curves (data not shown) determined that 2µg/mL puromycin was sufficient to kill all unmodified cells in 3 days. Therefore, 2µg/mL of puromycin was used for 3 days to selected cells with the integrated lentiviral constructs, and these cell lines were also passaged in 2µg/mL of puromycin for 3 days each time after recovery from -80°C storage.

The pL-EGFPN1 construct (discussed later) provided HCC cell lines with G418 resistance. Kill curves (data not shown) determined that 800µg/mL G418 was sufficient to kill all unmodified cells after 10 days. Therefore, 800µg/mL of G418 was used for 10 days to select HCC cell lines modified with this construct.

2.3 Prokaryotic Culture

2.3.1 Bacterial Strains

Table 2.3: Bacterial strains used for plasmid cloning and protein expression.

Strain	Use	Antibiotic Resistance
KRX	Protein Expression	
BL21(DE3) RIPL	Protein Expression	50µg/mL Chloramphenicol
Origami B (DE3)	Protein Expression	
DH5α	Cloning	

2.3.2 Growth Media

Bacteria were grown on Luria and Burrows agar (LB agar, 5g/L yeast extract, 10g/L tryptone, 10g/L sodium chloride, 15g/L agar) supplemented with antibiotics (**Table 2.3**) and 1% glucose when necessary. For minipreps, single colonies were grown overnight in liquid Luria and Burrows broth (LB broth, 5g/L yeast extract, 10g/L tryptone, 10g/L sodium chloride) supplemented with necessary antibiotics Table 2.3).

For protein expression, single colonies were firstly grown overnight in LB broth with antibiotics and 1% glucose when necessary Table 2.3). Overnight culture was then used to inoculate a larger volume of Terrific Broth (TB, 24g/L yeast extract, 20g/L tryptone, 0.4% v/v glycerol, 17mM KH_2PO_4, 72mM K_2HPO_4) supplemented with necessary antibiotics Table 2.3).

2.4 Antibodies

Table 2.4: Western blot antibodies. TBST, Tris-Buffered Saline Tween20; BSA, Bovine Serum Albumin

Antibody	Dilution	Buffer	Secondary antibody
Anti-Hormad2-rabbit (Invitrogen, PA5-43944)	1:500	5% skim-milk/TBST	Anti-rabbit (Cell Signaling, 7074S) 1:1000
Anti-ATM-mouse (Invitrogen, MA5-31599)	1:100	5% skim-milk/TBST	Anti-mouse (Cell Signaling, 7076) 1:1000
Anti-phosphoATM-mouse (Biolegend, 651202)	1:500	5% BSA/TBST	Anti-mouse (Cell Signaling, 7076) 1:1000
Anti-ATR-rabbit (Invitrogen, PA5-85507)	1:1000	5% skim-milk/TBST	Anti-rabbit (Cell Signaling, 7074S) 1:1000
Anti-phosphoATR-rabbit	1:1000	5% BSA/TBST	Anti-rabbit (Cell Signaling, 7074S) 1:1000
Anti-phospho-p53-rabbit	1:1000	5% BSA/TBST	Anti-rabbit (Cell Signaling, 7074S) 1:1000
Anti-β-actin-mouse (Sigma Aldrich, A2228)	1:10000	5% skim-milk/TBST	Anti-mouse (Cell Signaling, 7076) 1:1000
Anti-β-Tubulin (Developmental Studies Hybridoma Bank, E7-c)	1:1000	5% skim-milk/TBST	Anti-mouse (Cell Signaling, 7076) 1:1000
Anti-PCNA-mouse (Sigma Aldrich, P8825)	1:3000	5% skim-milk/TBST	Anti-mouse (Cell Signaling, 7076) 1:1000

2.5 Primers

Table 2.5: Sequencing and cloning primers.

Name	Use	Sequence
T7 Forward	Identification	TACGACTCACTATAGGGAGACC
PCW57.1 Forward	Cloning HORMAD2 into PCW57.1 from pL-HORMAD2-FLAG	AAAGCTAGCGCCACCATGGCCACTGCTCA
HORMAD2 Forward 1	Identification	GGTTGTTTTGATGCTTTGG
HORMAD2 Reverse 1	Identification	GTAGTGCTGTTCTCCCGAAA
PCW57.1 Reverse (With Stop)	Cloning HORMAD2 into PCW57.1 from pL-HORMAD2-FLAG	AAAACCGGTTCAATCAATCATTTTCTGTTAGGGATGAAGACCTTCACTGG
PCW57.1 Reverse (No Stop)	Cloning HORMAD2 into PCW57.1 from pL-HORMAD2-FLAG	AAAACCGGTCCTTTTCTGTTAGGGATGAAGACCTTCACTGG
GFP Reverse	Identification	CATCTAATTCAACAAGAATTGGGACAACTC
CMV Forward	Identification	CGCAAATGGGCGGTAGGCGTG
LTR Reverse	Identification	CATTCCCCCCTTTTTCTGGAG
LNCX Forward	Identification/Sequencing	AGCTCGTTTAGTGAACCGTCAGATC
HORMAD2 Reverse 2	Identification/Sequencing	GCTCAAGGTCCTGCATCAG
O.PGK1b Reverse	Identification/Sequencing	GAACGGACGTGAAGAATGTG

2.6 Mammalian Gene Expression Constructs

2.6.1 Constitutive HORMAD2 Overexpression

pL-HORMAD2-FLAG was previously created by Miriam Wankell and, therefore, provided by the Hebbard Lab. pL-HORMAD2-uvGFP was created by subcloning the sequence from pIM013-HORMAD2-uvGFP with NotI-HF (R3189S, NEB) and ApaI (R636A, Promega) into the empty pLN1 plasmid, provided by Lionel Hebbard. Fragments were isolated by agarose gel purification and ligated using T4 DNA ligase (M0202S, NEB).

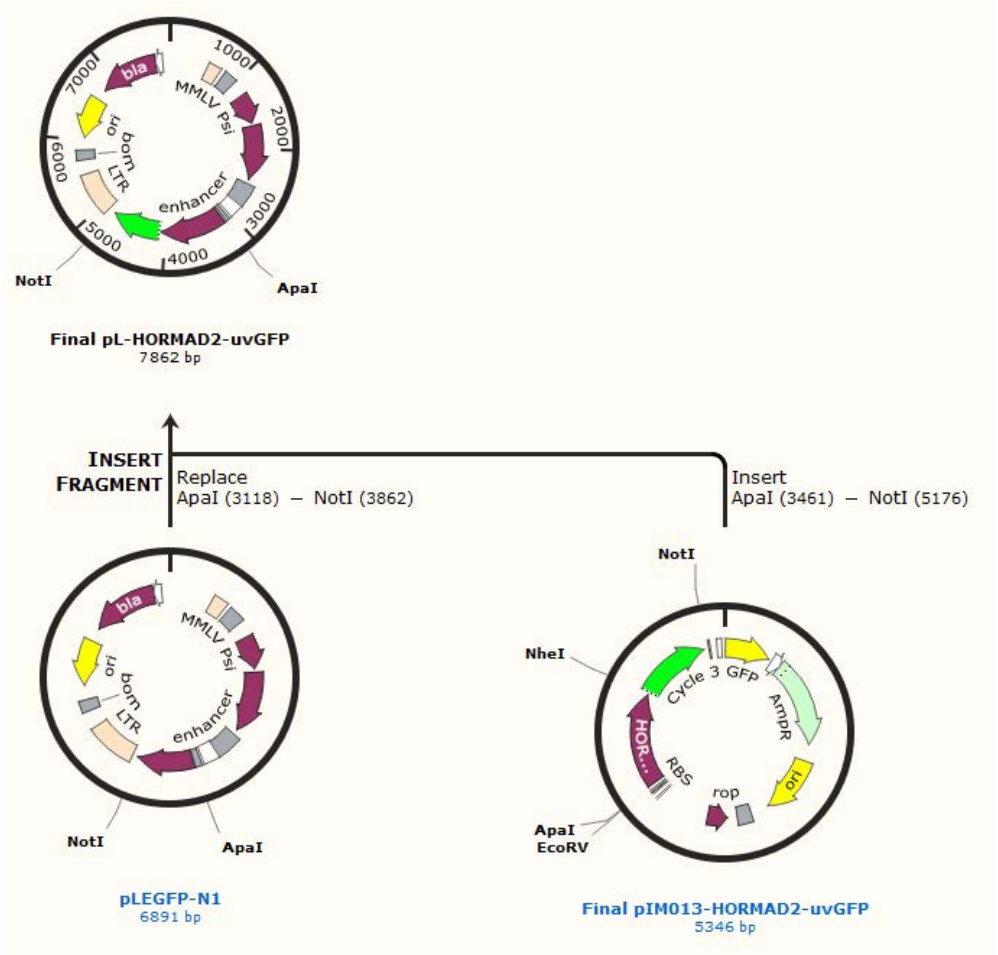

Figure 2.1: Cloning process for pL-HORMAD2-uvGFP. The HORMAD2-uvGFP sequence was subcloned from the pIM013-HORMAD2-uvGFP into the pLEGFP-N1 vector using ApaI and NotI restriction enzymes. AmpR, Ampicillin resistance; ori, origin; bla, Beta-Lactamase; MMLV Psi, Moloney Murine Leukemia Virus Packaging Signal; LTR, long terminal repeat; bom, basis of mobility; RBS, ribosome-binding sequence; rop, repressor of primer

2.6.2 Inducible HORMAD2 Overexpression

An agar stab of *E. coli* containing the tetracycline-inducible transgene expression plasmid PCW57.1 (#41393, Addgene) was purchased. Glycerol stocks were prepared after overnight culture in LB broth, and the plasmid was recovered using standard alkaline lysis miniprep techniques. A PCW57.1-HORMAD2-6HIS construct was created as described below by replacing the toxic ccdB insert in the plasmid with the HORMAD2 sequence. Additionally, a PCW(Empty) vector was created as a control by removing the ccdB insert.

The HORMAD2 sequence was amplified from the pL-HORMAD2-FLAG plasmid using the following primers: PCW57.1 Forward, PCW57.1 Reverse (With Stop), PCW57.1 Reverse (No Stop) (**Table 2.5**) – and a high fidelity Phusion Polymerase (M0530S, NEB). This PCR product as well as the PCW57.1 plasmid were digested using standard restriction digest with AgeI-HF (R3552S, NEB) and NheI-HF (R3131S, NEB). Some of the digested PCW57.1 was blunted using Klenow Large Fragment (M220A, Promega) for the PCW(Empty) construct, while the rest was dephosphorylated using Shrimp Alkaline Phosphatase (M820A, Promega). Digested fragments were purified using agarose gel purification. Finally, the fragments were re-ligated using T4 DNA ligase (M0202S, NEB).

The generated constructs were then amplified using standard miniprep techniques and stored in DH5α *E. coli* cells in glycerol stocks. The sequences were verified by PCR, restriction digest, and Sangar sequencing by AGRF.

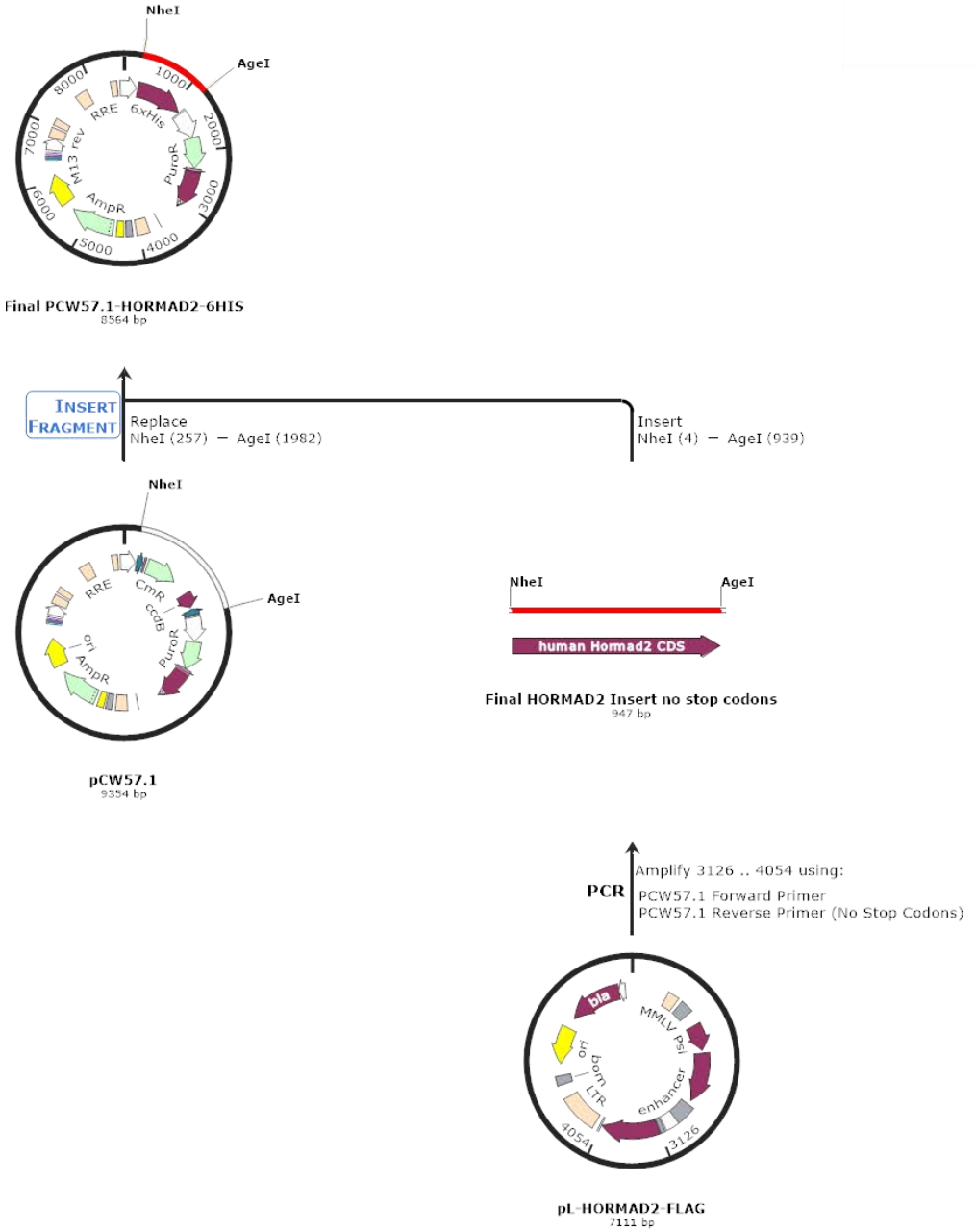

Figure 2.2: Cloning process for PCW57.1-HORMAD2-6HIS. The HORMAD2 sequence was amplified from the pL-HORMAD2-FLAG plasmid via polymerase-chain-reaction with the introduction of Nhe1 and Age1 restriction sites to allow integration into the PCW57.1 plasmid. AmpR, Ampicillin resistance; RRE, Rev response element; ori, origin; CmR, Chloramphenicol resistance; cccdB, Toxin CcdB; PuroR, Puromycin resistance; bla, Beta-Lactamase; MMLV Psi, Moloney Murine Leukemia Virus Packaging Signal; LTR, long terminal repeat; bom, basis of mobility.

2.6.3 shRNA Knockdown

Glycerol stocks of human HORMAD2 MISSION shRNA plasmids (Sigma Aldrich, SHCLNG-NM_152510, **Figure 2.3**) were purchased for 5 targeted regions of HORMAD2 (

Table 2.6). A non-targeting shRNA plasmid (glycerol stock provided by Lionel Hebbard) was used for a negative control.

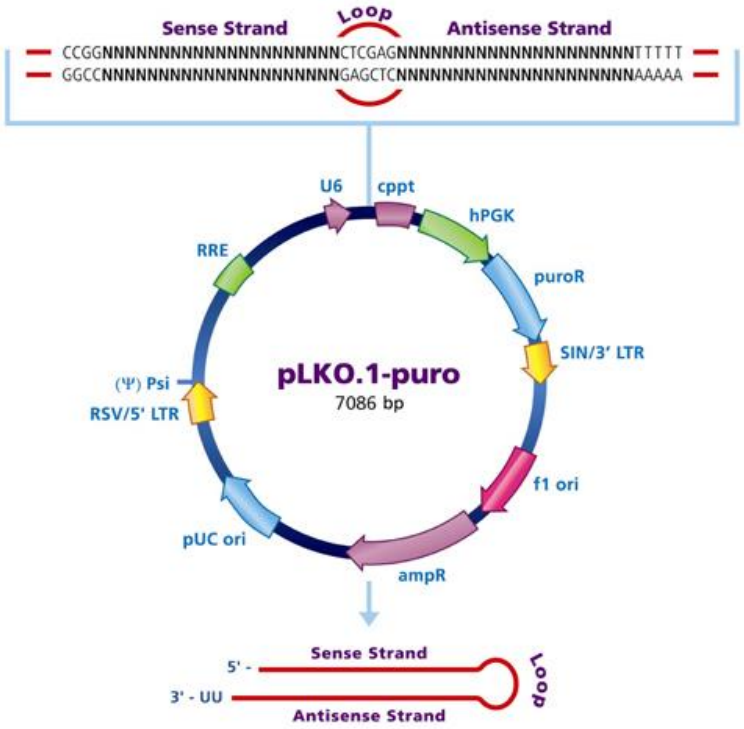

Figure 2.3: Sigma Aldrich MISSION shRNA vector map. U6, U6 Promoter; cppt, central polypurine tract; hPGK, human phosphoglycerate kinase eukaryotic promoter; puroR puromycin resistance gene for mammalian selection; SIN/3′ LTR 3', self-inactivating long terminal repeat; f1 ori, f1 origin of replication; ampR, ampicillin resistance gene for bacterial selection; pUC ori, pUC origin of replication; 5' LTR, 5' long terminal repeat; Psi RNA packaging signal; RRE, rev response element

Table 2.6: Human Hormad2 shRNA sequences. 3'UTR, 3' untranslated region; CDS, coding sequence

Plasmid ID	Clone ID	Region	Sequence
sh-HORMAD2 # 1	NM_152510.2-1452s21c1	3'UTR	CCGGGCTTATCCAGGTGTGATTTATCTCGAGATAAATCACACCTGGATAAGCTTTTTTG
sh-HORMAD2 # 2	NM_152510.1-912s1c1	CDS	CCGGGCAGTCAGCAAAGTTCTGAGTCTCGAGACTCAGAACTTTGCTGACTGCTTTTTTG
sh-HORMAD2 # 3	NM_152510.2-1173s21c1	3'UTR	CCGGCATAAGGAAAGCGGGTCAATACTCGAGTATTGACCCGCTTTCCTTATGTTTTTTG
sh-HORMAD2 # 4	NM_152510.2-1394s21c1	3'UTR	CCGGGAATCCCTCTCTAGCTGTATTCTCGAGAATACAGCTAGAGAGGGATTCTTTTTTG
sh-HORMAD2 # 5	NM_152510.2-295s21c1	CDS	CCGGGTCCCGGGTCACTGCATATTACTCGAGTAATATGCAGTGACCCGGGACTTTTTTG

2.7 Prokaryotic Recombinant Gene Expression Construct

A plasmid for bacterial recombinant gene expression was designed and purchased from Bioneer. The HORMAD2 coding sequence (CDS) was retrieved from the Kyoto Encyclopedia of Genes and Genomes (KEGG) and checked for rare codon-usage using Bioline Rare Codon Search. The HORMAD2 CDS was modified slightly to alter the DNA sequence for three Arginine positions (Arg65, Arg305, Arg308, agg>cgc) which utilized rare codons from *E. coli*. These were silent mutations that did not alter the HORMAD2 amino acid sequence.

The codon-optimised HORMAD2 sequence was inserted into the pIM013-pETuvGFP plasmid in frame with the N-terminal 6HIS tag and C-terminal GFP fusion-tag. This plasmid additionally provides ampicillin resistance, and is controlled by a T7 polymerase promoter.

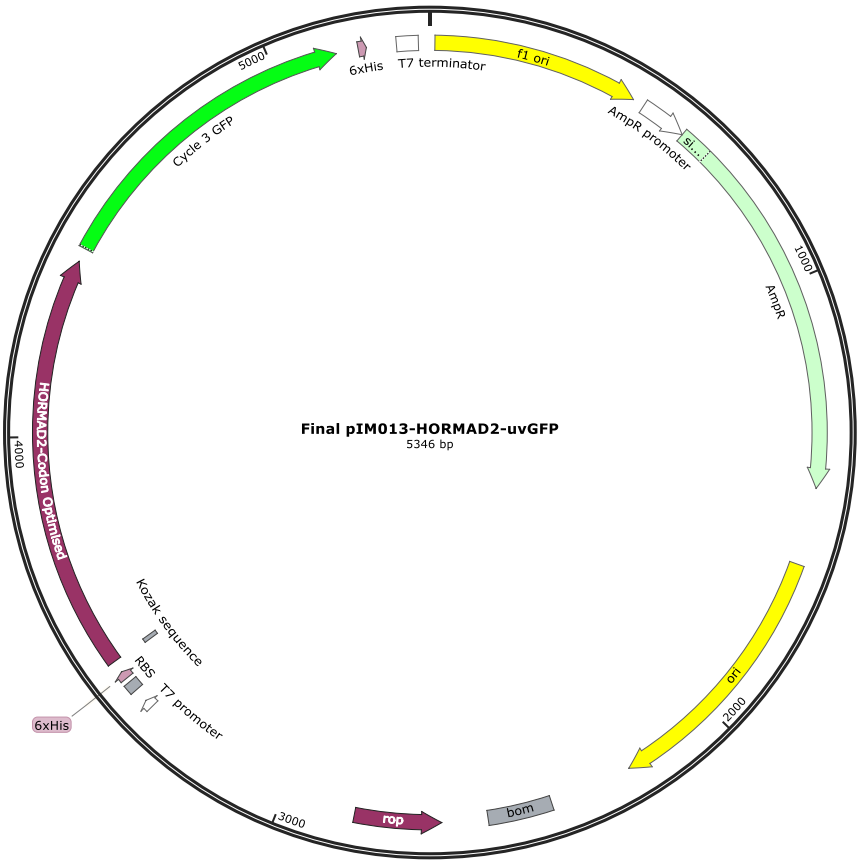

Figure 2.4: Plasmid map of pIM013-HORMAD2-uvGFP. Created from the pIM013-pETuvGFP vector, a modified pET vector provided by Patrick Schaeffer. 6xHis, 6-Histidine Tag; AmpR, Ampicillin Resistance; bom, basis of mobility; rop, repressor of primer; RBS, ribosome-binding sequence

2.8 Plasmid Recovery

Glycerol stocks and agar stabs were cultured according to manufacturer's protocols. In brief, 50 µL of glycerol stocks were mixed with 500 µL of TB without antibiotics and incubated at 37°C for 20 minutes, then these cultures were streaked onto antibiotics-containing LB agar plates. Agar stabs were directly streaked onto LB agar plates. Plates were incubated at 37°C for 18hrs or until single colonies were visible. Single colonies were selected for minipreps.

Plasmids for mammalian cell culture were prepared using either GenElute Endotoxin-Free Midiprep (PLED35-IKT, Sigma Aldrich) kits as per the manufacturers guidelines or using Alkaline lysis with phenol:chloroform endotoxin removal and sterile filtering.

Promega Wizard Plus SV Miniprep (A1330, Promega) Kits or alkaline lysis were used for non-mammalian cell culture experiments.

2.8.1.1 Alkaline Lysis

DH5α cells containing plasmids were cultured overnight in 2mL of LB broth with antibiotics at 37°C with 200 rpm shaking. 1.5mL of overnight culture was pelleted at 35 000 x g for 2 minutes, and the

supernatant was removed. Bacterial pellets were suspended in 100 µL of Alkaline Lysis Solution I (50 mM glucose, 25 mM Tris-Cl pH 8.0, 10 mM EDTA pH 8.0, 20 µg/mL RNase A). 200 µL of Alkaline Lysis Solution II (0.2 N NaOH, 1% w/v SDS) was added to each suspension, which were then inverted gently 6 times, and stored on ice. 150 µL of ice-cold Alkaline Lysis Solution III (60% 5 M potassium acetate, 11.5% glacial acetic acid) was added and tubes inverted several times. Tubes were stored on ice for 5 minutes and centrifuged at 35 000 x g for 5 minutes at 4°C. The supernatant was transferred to a fresh tube and an equal volume of phenol was added, vortexed, and centrifuged at 35 000 x g for 2 minutes at 4°C. The upper layer was collected and transferred to a new tube, and the above step was repeated with chloroform. Nucleic acids were precipitated by adding 2 volumes of 100% ethanol. The solutions were vortexed and centrifuged at 35 000 x g for 5 minutes at 4°C. Supernatant was removed, 1 mL of 70% ethanol was added, and centrifuged at maximum speed for 5 minutes at 4°C. Supernatant was removed, and the nucleic acid pellet was resuspended in 100 µL of TE buffer. Plasmid yield was determined via NanoDrop and by observing migration on agarose gel.

2.9 Lentivirus Production

Glycerol stocks for the packaging plasmid, pCMVΔR8.2 (addgene, #12263), and VSV-G expressing envelope plasmid, pMD2.G (addgene #12259) were supplied by Lionel Hebbard. These plasmids were recovered as outlined previously.

HEK293T cells were transfected with the plasmids using a modified calcium chloride transduction protocol (Green Trono Lab). In brief, cells of 70-90% confluency were split into three 10cm plates per viral construct the day before transduction. For each plate of cells to be transduced, 20 µg of shRNA or transgene expression plasmid, 15 µg of packaging plasmid (pCMVΔR8.2), and 5 µg of envelope plasmid (pMD.G) were mixed to a final volume of 250 µL in HEPES buffered water (2.5 mM pH 7.3) in a 1.5 mL Eppendorf tube. Then, 250 µL of 0.5M $CaCl_2$ was added dropwise to the tube. The mixture was added to 500 µL of 2 x HeBS dropwise in a 50 mL falcon tube with constant mechanical pipetting to bubble the HeBS. Mixtures were incubated at room temperature for 20-30 minutes. The mixture was added to HEK293T cells in a spiral movement, followed by gentle shaking, and the cells were incubated for 24 hours. After 24 hours, the media was replaced with fresh media and cells were incubated for another 24 hours. Virus was collected by harvesting the media every 24 hours for 2-3 days. Harvested media was filtered through 0.45 µm filter, and either frozen at -80°C or centrifuged in Beckman Ultracentrifugation tubes at 21,000 rpm for 1.5 hours at 4°C, to concentrate the virus. Supernatant was gently decanted, and the concentrated viral pellet was resuspended in 100 µL of ice-cold PBS and stored at -80°C.

2.10 Lentiviral Transduction into HCC Cell Lines

For lentivirus transduction, 1 μL of each concentrated virus was diluted in Opti-MEM reduced-serum media (Gibco, 31985062) in a 24-well plate in duplicate. Virus was left to incubate in the wells for approximately 20 minutes. Huh7 or Hep3B cells were seeded as 10,000 cells per well on top of the virus or on blank wells for controls. After 72 hours, media was removed and replaced with fresh DMEM media containing 10% FBS and 2 μg/mL Puromycin to select for transduced cells. Cells were left to select for 3 days and positively transduced cells were grown and harvested for experimentation. The cell lines were termed as the following: sh-Scram (negative control transduced with non-targeting lentivirus) sh-HORMAD2 #1-5 (transduced with HORMAD2 shRNA lentivirus #1-5), PCW(Empty) (negative control transduced with empty PCW57.1 lentivirus), PCW-HORMAD2 (transduced with PCW57.1-HORMAD2 lentivirus) and PCW-HORMAD2-6HIS (transduced with PCW57.1-HORMAD2-6HIS lentivirus). Normal cells (non-transduced) were used as an additional control. To check for successful viral integration, protein lysates were obtained in the absence of puromycin and Western Blots were performed with a Hormad2 and/or 6HIS antibody.

2.11 Protein Lysates

2.11.1 Standard Protein Lysates

All protein lysates for immunoblotting were prepared using radio-immunoprecipitation assay (RIPA) buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% or 0.5% SDS) (Note: SDS concentration increased from 0.1% to improve the yield of nuclear proteins) with freshly added protease inhibitors (2 mM sodium orthovanadate, 50 mM NaF, 1 mM sodium molybdate, 40 mM β-glycero-phosphate, 1 mM PMSF, protease inhibitor cocktail (Sigma, P8340)). Adherent cells were washed twice with ice-cold PBS, before RIPA buffer was added. Plates were rocked on ice for 5 minutes, then scraped down into 1.5 mL centrifuge tubes. 5 μL of DNase 1 (NEB, M0303S) per 100uL of lysate was added. Lysates were incubated on ice for 30 minutes with intermittent vortexing, then passed through a 29G syringe until clear. Cell debris was pelleted at 18 000 x g for 10 minutes, 4°C, and the supernatant stored at -80°C.

2.11.2 Nuclear Fractionation

Hep3B PCW-HORMAD2-6HIS cells were grown on 10cm cell culture plates to 80% confluency and HORMAD2-6HIS protein expression was induced with 2μg/mL doxycycline. Cells were washed with ice-cold PBS then covered with 5mL of ice-cold hypotonic lysis buffer (10 mM Tris-HCl pH 8, 0.1 mM DTT, 1 mM PMSF). The plate was incubated on ice with rocking for 10 minutes. Cells were scraped from the plate with a rubber policeman and collected into a glass 5 mL douncer. The cell suspension was dounced 10 times with minimal force applied to lyse the cells while keeping the nuclei intact. The lysate was transferred to a 15 mL Falcon tube and centrifuged at 340 x g for 10 minutes at 4°C to pellet the cell nuclei.

The supernatant was collected as the cytoplasmic fraction with insoluble membranous fraction remaining. Firstly, the insoluble fraction was pelleted via ultracentrifugation (Thermo Scientific, MTX 150) at 35000 rpm for 30 minutes at 4°C then resuspended in 50uL of RIPA buffer (50 mM Tris pH 7.5, 150mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1% SDS, 2mM sodium orthovanadate, 50 mM NaF, 1 mM sodium molybdate, 40 mM β-glycero-phosphate, 1 mM PMSF, protease inhibitor cocktail (Sigma, P8340)) and centrifuged again at 18000 rpm, 10 minutes, 4°C to produce the membrane protein fraction. Secondly, the remaining supernatant was precipitated by adding 313mg of ammonium sulphate per mL of lysate then ultracentrifuged at 35000 rpm, 30 minutes, 4°C and the precipitated protein pellet was resuspended in 75 µL of RIPA buffer.

The pelleted nuclei from the initial low-speed centrifugation step were resuspended in 50 µL of RIPA buffer and clarified by passing through an insulin syringe several times and digesting with 1uL DNase1 for 30 minutes on ice with intermittent vortexing. Remaining cellular debris was removed by centrifugation at 18000 rpm, 10 minutes, and 4°C.

2.11.3 Protein Quantification

All protein lysates were quantified in triplicate using the DC Protein Assay (Bio-Rad, 5000112). A standard curve was prepared using 0.2-3 µg/uL BSA in RIPA buffer in triplicate.

2.12 Western Blots

Once quantified, 20 or 30ug aliquots were prepared with 5x Laemmli SDS sample buffer (10% Glycerol, 60 mM Tris HCl pH 6.8, 2% SDS, 0.01% bromophenol blue) and stored at -80. Prior to loading, fresh dithiothreitol (DTT) was added to 100mM, and the samples were heated to 95°C for 10 minutes. Samples were loaded into 4-20% TruPAGE Precast Gels (Sigma, PCG2012-10EA) or homemade polyacrylamide gels at specific concentrations. Samples were loaded into polyacrylamide gel along with 11-190 kDa Prestained Protein Marker (Cell Signaling, #13953). Samples were run through the gel in running buffer (60 mM Triethanolamine, 40 mM Tricine, 0.1% SDS) at 60 V until through the stacking gel, then at 100 V until the dye front reached the bottom of the gel.

Once run, a transfer stack was prepared to transfer the separated protein onto a PVDF membrane (activated with methanol for 10 seconds, rinsed in H_2O, and equilibrated in transfer buffer for 20 minutes). Protein transfers were performed overnight at 30 V, 4°C or for 1-1.5hrs at 100 V, 4°C in ice bath in transfer buffer (12.5 mM Trizma base, 96 mM Glycine) with constant circulation of buffer using a magnetic stirrer.

After transfer, the membrane was cut if necessary using a clean scalpel, then immediately blocked in 5% skim milk in TBST for non-phosphorylated proteins, or 5% BSA in TBST for phosphorylated proteins for at 30 minutes to 6 hours and room temperature with rocking.

After blocking, membranes were probed with primary antibodies diluted in blocking solution overnight at 4°C with rocking. Membranes were washed 3 times for 10 minutes with TBST. Membranes were incubated with secondary, HRP-linked antibodies for 45-60 minutes at room temperature. Membranes were washed 5 times for 5 minutes with TBST. Membranes were incubated with 0.5-1mL of freshly prepared ECL reagent (100 mM Tris-HCl pH 8.8, 1.25 mM Luminol, 2 mM 4IPBA, 0.0162% H_2O_2) for 1-3 minutes, then imaged (Syngene G:BOX Chemi XRQ).

2.13 Cell Proliferation Analysis

2.13.1 Holomonitor

Cell proliferation was determined using a Holomonitor (Phi) microscope in Huh7, Huh7 sh-Scram, Huh7 sh-HORMAD2 knockdown, Huh7 PCW(Empty) and Huh7 PCW57.1-HORMAD2-6HIS cells. Cells were plated in 6-well plates at 75 000 cells per well, and PCW57.1 cells were cultured in doxycycline for 24 hours before the experiment. Cells were cultured under the Holomonitor for 96 hours and cell count was analysed every hour at 10 different positions.

2.13.2 BrdU Assay

Cell proliferation was also evaluated using a Bromodeoxyuridine (BrdU) Colorimetric assay (Roche), according to the manufacturer's protocol, with some adaptations as follows. Huh7, Huh7 sh-Scram, and Huh7 sh-HORMAD2 knockdown (n=8) cells were cultured in 96-well plates to evaluate proliferation after 24, 48, and 72 hours. Cells were incubated with BrdU (Roche kit) for 2 hours, and fixed with the fix-denaturant solution provided from the kit as per instructions. Cells were washed once with PBST (PBS with 0.1% Tween20), and incubated with anti-BrdU-mouse antibody (diluted 1:100 in 1% BSA/PBST) (antibody provided by Developmental Studies Hybridoma Bank, DSHB) for one hour. The antibody solution was removed, cells were washed three times with PBST, and incubated with anti-mouse-HRP antibody (diluted 1:1000 in 1% BSA/PBST) for 30 minutes. The antibody solution was removed, cells were washed three times with PBST, and 100 µL of TMB substrate solution (421101, BioLegend) was added to each well and incubated for approximately 25 minutes. The reaction was stopped with 25 µL per well of. 1M H_2SO_4 and the absorbance was measured at 450 nm on a FLUOstar Omega microplate reader (BMG Labtech).

2.14 MTT Cell Viability Assays

Cells were seeded in 96-well plates as 10 000 cells per well in replicates of 4-8 for each treatment depending on space availability in each experiment, and allowed to adhere overnight. Media was replaced with fresh media containing (concentrations varied depending on sensitivity of cell lines, see individual figures for exact concentrations):

- 0-128 μM Sorafenib (Adooq Biosciences, A10001, 5 mM stock solution in DMSO) in DMEM + 10% FBS
- 0-128 μM Cisplatin (Selleckchem, S1166, 10 mM stock solution in saline) in DMEM + 10% FBS
- 0-1000 μM H_2O_2 in FBS-free DMEM
- 5 mM Glucose or Fructose in glucose-free DMEM (11966-025, Gibco)

Cells were left for 24-48 hours and cell viability was determined with MTT (3-(4,5-Dimethylthiazol-2-yl)-2,5-Diphenyltetrazolium Bromide). In brief, MTT (Sigma Aldrich, M5655, stock solution 5 mg/mL in sterile dH2O) was added to each well to a final concentration of 0.5 mg/mL and left to incubate at 37°C for approximately 2 hours. Media was aspirated and 50 μL of DMSO was added to each well and incubated further at 37°C for 15-20 minutes, or until the colour was evenly distributed. Absorbance was read at 540 nm on a FLUOstar Omega microplate reader (BMG Labtech).

2.15 LINAC Radiation to Induced DNA Damage

Cells were grown in T75 flasks (Greiner) to approximately 80% confluency. Cells were irradiated using 6mV photons using an Elekta linear accelerator. Radiation was delivered in fractions of 1.8 Gy. For higher radiation doses, 1.8Gy doses were delivered immediately after, to produce low dose (1.8 Gy), medium dose (1.8 x 3 = 5.4Gy), and high dose (1.8 x 5 = 9Gy) radiation. After radiation, cells were moved to the laboratory as fast as possible and trypsinised (20±5 minutes) and put on ice (31±3 minutes) and protein lysates were prepared (50±5 minutes) (Average ± SD).

2.16 Tumoursphere Assay

After radiation-treatment, cells were trypsinised to detach them from the cell culture flasks and counted. Cells were seeded at 3.0 x 10^4 cells per well in 6-well low attachment plates (Sigma Aldrich) in triplicate (n=3) in sphere assay media consisting of 2% B27 Supplement (Gibco), 20 ng/mL hEGF (Sigma Aldrich), 10 ng/mL basic fibroblast growth factor (bFGF; Gibco), 2 μg/mL heparin (Sigma Aldrich), 5 μg/mL insulin (Gibco), and 0.5 μg/mL hydrocortisone. Spheres were allowed to form for seven days of incubation at 37°C in 5% CO_2, after which they were counted manually under a microscope.

2.17 Colony-Formation Assay

Non-irradiated cells were seeded as 1000 cells per well in a 6-well plate (n=5-8). After radiation-treatment, cells were seeded as 2000, 4000, 8000 cells per well in a 6-well plate. Colonies were left to grow for 14 days and colonies of minimum 50 cells were counted. Due to the need to move radiated cells between buildings, some samples were lost due to environmental contamination, thus the sample sizes for each group will be listed. For 1.8Gy radiation, seeding of 2000 and 4000 cells were used for the data presented (n=5-6). For 5.4Gy radiation, seeding of 4000 and 8000 cells were used (n=1-3). The

plating efficiency (PE) and surviving fraction (SF) was calculated based on previously established protocol (211). In brief, PE was calculated for the unirradiated cells with the following equation:

$$PE = \frac{no.of\ colonies\ formed}{no.of\ cells\ seeded}\ x\ 100\%$$

The SF relates to the number of colonies formed in response to irradiation treatment relative to the PE, and was calculated as follows:

$$SF = \frac{no.of\ colonies\ formed\ after\ treatment}{no.of\ cells\ seeded\ x\ PE}$$

2.18 Alkaline Comet Assay

Alkaline comet assays were performed based on previously described protocols (212, 213). Briefly, untreated and DNA-damaged cells were resuspended in cold PBS at 500 000 cells/mL. 100 µL of cell suspension was mixed with 1mL of 1% molten low melting-point agarose, mixed gently, and 30 µL of this suspension was added to a dried glass slide that had been coated with molten 1% normal melting-point agarose. A coverslip was placed over the cells to spread the mixture, and the agarose was allowed to set for 10-30 minutes at 4°C. The coverslips were removed and slides were submerged in a covered dish containing alkaline lysis solution (2.5 M NaCl, 0.1 M Na_2EDTA, 0.01 M Tris, 0.2 M NaOH, 1% Triton-X-100, pH 10) and incubated overnight at 4°C in the dark. Slides were briefly rinsed in $_{dd}H_2O$, and submerged in electrophoresis buffer (0.2 M NaOH, 0.001 M Na_2EDTA, pH 13) to allow the DNA to unwind for 1 hour at 4°C in the dark. Slides were placed in an electrophoresis chamber, covered with electrophoresis buffer, and run at 1V/cm in 4°C for 30 minutes. Slides were immersed in dH2O twice for 5 minutes, dehydrated in 100% ethanol for 5 minutes, and dried at 37°C for 15 minutes in the dark. Next, 100 µL of DNA stain solution (0.01 M Tris-HCl pH 7.4, 0.001 M EDTA, 1x GelRed) was added onto each slide and stained for 15 minutes at room temperature in the dark. Slides were rinsed briefly in dH2O and dried at 37°C in the dark. Slides were imaged with an Olympus BX51 Fluorescence Micrscope using an Olympus U-RFL-T Fluorescent Lamp and Olympus DP70 camera system. Approximately 50 comets per sample were analysed manually using Tritek CometScore v2.0.0.38. For the H_2O_2 assay, 47-54 comets were counted (n=47-54). For the radiation assay, 44-51 comets were counted (n=44-51).

2.19 Recombinant HORMAD2-uvGFP Expression

2.19.1 Transformation

20 µL of competent cells were combined with 10-100 ng of plasmid DNA and incubated on ice for 20 minutes. After incubation, the cells were heated to 42°C for 30 seconds to facilitate uptake of the plasmid, and were immediately placed back on ice for 2 minutes. 500 µL of LB medium was added to

the transformed cells, and the cell suspension was then incubated at 37°C for 1 hour with shaking. 100-500 µL of this culture (dependent on the plasmid used) was plated on LB agar containing appropriate antibiotics and glucose if necessary, and incubated at 37°C overnight.

2.19.2 Protein Expression

Single colonies of cells on LB agar plates were selected and grown overnight in 5 mL of LB broth at 37°C, 250 rpm with appropriate antibiotics and 1% w/v glucose for KRX cells. Overnight cultures were used to inoculate 100 mL of appropriate media (KRX, TB; BL21(DE3)RIPL and Origami B (DE3), Autoinduction Media (TB, 4 mM Glucose, 0.4 mM Galactose)) to an OD_{600} of approximately 0.125, and grown at 37°C, 250rpm until OD_{600} ~ 1. For BL21(DE3)RIPL and Origami B (DE3) induction was automatic, while for KRX expression was induced with the addition of rhamnose to 0.1% v/v final. In either case, expression was performed at 16°C, 250 rpm for approximately 72 hours. After protein expression, bacteria were pelleted at 4800 rpm for 30 minutes at 4°C, snap-frozen in liquid nitrogen and stored at -80°C.

2.19.3 Ni-IMAC Protein Purification

Bacterial pellets were thawed and resuspended in phosphate lysis buffer (45 mM Na_2HPO_4, 5 mM NaH_2PO_4, pH 7.8, 300 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol, 10% glycerol) to 7.5 mL per 1g of bacterial mass. After thorough resuspension, cells were lysed with two passes of a French press at 700 psi, and the cell debris was removed via centrifugation at 40 000 x g for 30 minutes at 4°C. The clarified lysate was passed through a bed of Ni-NTA resin in a column, washed three times with 5 mL lysis buffer and eluted with multiple volumes of 1.5 mL elution buffer (45 mM Na_2HPO_4, 5 mM NaH_2PO_4, pH7.8, 300 mM NaCl, 200 mM imidazole, 2 mM β-mercaptoethanol, 10% glycerol). Successful elutions were combined with equal volumes of saturated ammonium sulphate to precipitate protein overnight at 4°C, then collected via centrifugation at 18000 rpm for 30 minutes and resuspended in phosphate buffer (45 mM Na_2HPO_4, 5 mM NaH_2PO_4, pH 7.8, 2 mM β-mercaptoethanol, 10% glycerol).

Purification yield was evaluated by running small aliquots on a 12% acrylamide SDS-PAGE (Resolving gel: 375 mM Tris pH 8.8, 12% v/v 29:11 acrylamide/bis-acrylamide, 0.1% v/v SDS, 0.1% v/v ammonium persulphate, 0.1% v/v tetramethylethylenediamine; Stacking gel: 140 mM Tris pH 6.8, 5.575% v/v 29:11 acrylamide/bis-acrylamide, 0.1% v/v SDS, 0.1% v/v ammonium persulphate, 0.1% v/v tetramethylethylenediamine). GFP fluorescence was measured using a blue light and 525 filter before Coomassie Brilliant Blue R-250 staining to evaluate total protein.

2.20 DSF-GTP

Differential scanning fluorometry of a GFP-tagged protein (DSF-GTP) was used to identify the melting temperature of HORMAD2-uvGFP. DSF-GTP was tested using different buffers by preparing aliquots of approximately 10ug of recombinant HORMAD2-uvGFP to 20uL volumes in various buffers (50 mM Phosphate Buffer pH 7.8, 50 mM Tris-HCl pH 7.5, 50 mM HEPES pH 7.4, 50 mM Citrate Buffer pH 6.0). Additionally, DSF-GTP was tested with potential interacting metal ions by preparing 20 µL aliquots of 10 ug HORMAD2-uvGFP in 50 mM HEPES pH 7.4 with 100 µM MgCl and 10 uM MnCl or 100 µM KAc and 10 µM ZnAc.

These aliquots were gradually heated in a real-time thermal cycler (BioRad, CFX96) from 25°C to 90°C in increments of 0.5°C every 30 seconds and the GFP fluorescence was read at each increment.

2.21 GFP-BASTA

For the GFP-Based Stability Assay (GFP-BASTA), recombinant HORMAD2-uvGFP was investigated alongside recombinant uvGFP as a control. Quantities of these two proteins were matched by fluorescence rather than gross protein mass, which is dependent on the purity of the protein preparation as well as any fluorescence quenching from the fusion protein, meaning that the protein masses were not equal. Aliquots of approximately 9ug of HORMAD2-uvGFP combined with approximately 2ug of uvGFP were combined to 20uL volumes in phosphate buffer (45 mM Na_2HPO_4, 5 mM NaH_2PO_4, 2 mM β-mercaptoethanol, 10% glycerol). Protein mixtures were heated over a temperature gradient for 10 minutes using a thermal cycler (BioRad MyCycler) cooled at 4°C for 10 minutes, and then centrifuged at 58 000 x g for 15 minutes. To detect a loss of protein due to aggregation, the fluorescence of the supernatant was measured using a fluorescence plate reader (BMG Labtech, FLUOstar Omega) and samples from the supernatant were run on an SDS-PAGE.

Once the aggregation temperature of HORMAD2-GFP was determined, 10 ug aliquots of HORMAD2-uvGFP were prepared under the following conditions:

- Control (No added molecules)
- 1 mM $MgCl_2$
- 5 µM ZnAc
- 100 µM $MnCl_2$
- 90 ng/µL genomic Bovine dsDNA restriction enzyme digested into short dsDNA
- DNA (as above) combined with 1 mM $MgCl_2$, 5 µM ZnAc, 100 µM $MnCl_2$
- 5 mM ATP

These samples were all heated the HORMAD2-uvGFP aggregation temperature for 10 minutes, cooled at 4°C for 10 minutes, and then centrifuged at 58000 x g for 15 minutes, and the remaining fluorescence was detected using a fluorescence plate reader (BMG Labtech, FLUOstar Omega).

2.22 GFP-EMSA

Samples were prepared using 6ug aliquots of HORMAD2-uvGFP or RNaseA (Sigma Aldrich, R4642) digested HORMAD2-uvGFP (treated with 0.02Kunitz for 1 hour at 25°C), 1ug of Hep3B genomic DNA (isolated using Promega Kit A1120), 60 ng of BioO (short <50 bp dsDNA), and 1.8 μg of restriction-enzyme digested bovine genomic DNA. Controls consisted of DNA samples only or recombinant protein only. All samples were prepared to equal volumes in phosphate buffer and then incubated at 25°C for 10 minutes before being run on a 1% agarose gel at 80 V for 30 minutes. GFP fluorescence was analysed with blue light and 525 filter, before the gel was post-stained with GelRed and analysed with UV light and 605M filter to capture DNA migration.

2.23 Statistics

All statistics were calculated using either GraphPad Prism 9 or the statistical tests within specific R packages where relevant and specified. For comparisons between groups, normality tests were performed using the D'Agostino-Pearson test. Normally distributed data was analysed using student T-tests and single-way ANOVA with Tukey's multiple comparisons test. Non-normally distributed data was analysed using the Mann-Whitney test and Kruskal-Wallis with Dunn's multiple comparisons test.

Chapter 3: *In Silico* Characterisation of HORMAD2 in Human HCC

3.1 Introduction

Although all known and proposed functions of HORMAD2 are restricted to germline cells, numerous lines of evidence suggest a role in the liver and in HCC. To elucidate potential functions and identify the significance of this role, a publicly available dataset from a large-scale human HCC sequencing project performed by the Cancer Genome Atlas (TCGA) was investigated. The TCGA cancer genomics program characterised 33 cancer types, and all data is publicly available for research use. Indeed, **Figure 3.1** shows that liver hepatocellular carcinoma (LIHC) and cholangiocarcinoma (CHOL), a rarer form of liver cancer, are the only two cancer types to show prominent HORMAD2 expression. Because the TCGA-CHOL cohort is far smaller (Normal n=9, Tumour n=13), the LIHC data was used for this analysis.

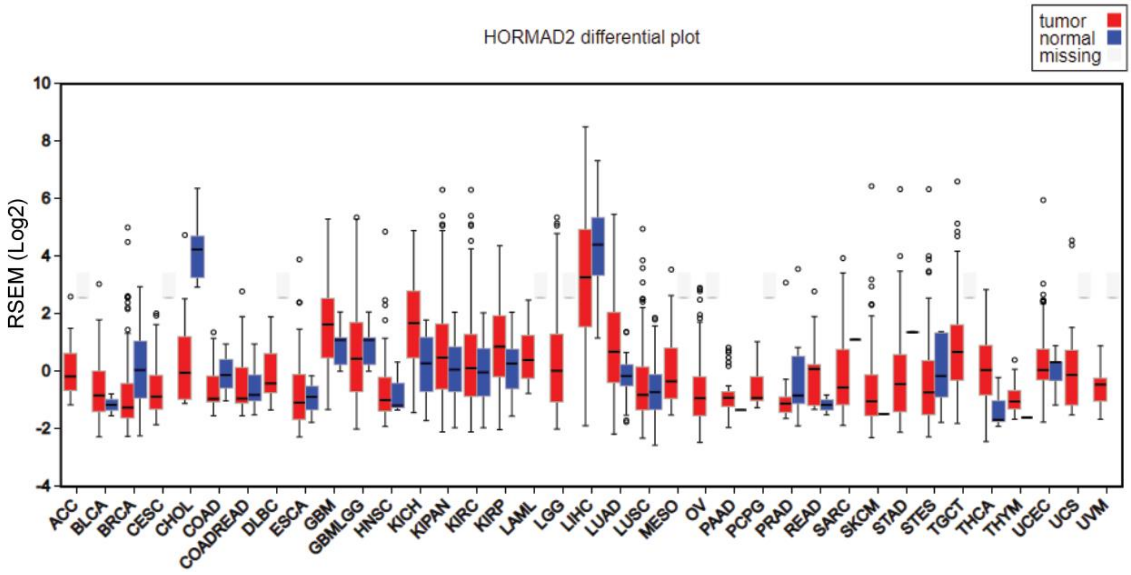

Figure 3.1: HORMAD2 expression across various cancer types. HORMAD2 is expressed most highly in hepatocellular carcinoma (LIHC) (Tumour n=286, Normal n=50) and cholangiocarcinoma (CHOL) (Tumour n=13, Normal n=9) Expression is shown as RSEM normalised read counts for both tumour and normal samples as indicated in the key. This graph was generated based on TCGA RNASeq data using FireBrowse (http://firebrowse.org/). ACC, Adrenocortical carcinoma; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; CNTL, Controls; COAD, Colon adenocarcinoma; DLBC, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; ESCA, Esophageal carcinoma; FPPP, FFPE Pilot Phase II; GBM, Glioblastoma multiforme; HNSC, Head and Neck squamous cell carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LAML, Acute Myeloid Leukemia; LCML, Chronic Myelogenous Leukemia; LGG, Brain Lower Grade Glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; MISC, Miscellaneous; OV, Ovarian serous cystadenocarcinoma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and Paraganglioma; PRAD, Prostate adenocarcinoma; READ, Rectum adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular Germ Cell Tumors; THCA, Thyroid carcinoma; THYM, Thymoma; UCEC, Uterine Corpus Endometrial Carcinoma; UCS, Uterine Carcinosarcoma; UVM, Uveal Melanoma; RSEM, RNA-Seq by Expectation-Maximisation; TCGA, The Cancer Genome Atlas

3.2 Results

3.2.1 Cohort Information

Table 3.1: Categorical descriptive data of the TCGA-LIHC cohort. Significance (p value) was calculated using a Chi Squared test, not including "Not Recorded" values. *, p < 0.05; **, p < 0.01; ***, p < 0.0001. BMI, Body mass index; HBV, Hepatitis B Virus; HCV, Hepatitis C Virus; NAFLD, Non-Alcoholic Fatty Liver Disease

			Entire Cohort (n=377)		Paired Tumour/Normal Cohort (n=50)	
Feature	**p**	**Type**	**Count**	**Percentage**	**Count**	**Percentage**
Adjacent Hepatic Tissue Inflammation	0.352	Mild	101	27%	11	22%
		None	119	32%	22	44%
		Severe	19	5%	2	4%
		Not Recorded	138	37%	15	30%
BMI Status	0.558	Normal	160	42%	14	28%
		Obese	68	18%	7	14%
		Overweight	92	24%	12	24%
		Underweight	21	6%	4	8%
		Not Recorded	36	10%	13	26%
Child Pugh Classification	**0.004 ****	A	223	59%	23	46%
		B	21	6%	9	18%
		C	1	0%	0	0%
		Not Recorded	132	35%	18	36%
Embolization Type	0.794	Chemoembolization	31	8%	7	14%
		Radioembolization	7	2%	2	4%
		Not Recorded	339	90%	41	82%
Fibrosis Ishak Score	**0.033 ***	0 - No Fibrosis	76	20%	22	44%
		1,2 - Portal Fibrosis	31	8%	4	8%
		3,4 - Fibrous Septa	30	8%	3	6%
		5 - Nodular Formation & Incomplete Cirrhosis	9	2%	1	2%
		6 - Established Cirrhosis	72	19%	5	10%
		Not Recorded	159	42%	15	30%
Gender	0.102	Female	122	32%	22	44%
		Male	255	68%	28	56%
History of Alcohol Consumption	0.053	No	259	69%	41	82%
		Yes	118	31%	9	18%
History of HBV or HCV	**0.018 ***	No	221	59%	38	76%
		Yes	156	41%	12	24%
History of NAFLD	0.773	No	358	95%	47	94%
		Yes	19	5%	3	6%
Pathologic Stage	0.830	Stage I	175	46%	18	36%
		Stage II	87	23%	11	22%
		Stage III	86	23%	12	24%
		Stage IV	5	1%	1	2%
		Not Recorded	24	6%	8	16%
Ethnicity	**8.75E-05 *****	American Indian or Alaska Native	2	1%	0	0%
		Asian	161	43%	6	12%
		Black or African American	17	5%	7	14%
		White	187	50%	34	68%
		Not Recorded	10	3%	3	6%
Treated with Radiation Therapy	0.856	No	243	64%	2	4%
		Yes	4	1%	0	0%
		Not Recorded	130	34%	48	96%
Relatives with Family History	**0.017 ***	No	212	56%	21	42%
		Yes	114	30%	24	48%
		Not Recorded	51	14%	5	10%

Table 3.2: Numerical descriptive data of the TCGA-LIHC cohort. Data availability indicates the percentage of patients with non-NA values in the clinical data sheet. Significance (p value) was calculated using a Student T-Test. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.0001$.

		Entire Cohort (n=377)		Paired Tumour/Normal Cohort (n=50)	
Clinical Data	**p**	**Mean ± SD**	**Availability**	**Mean ± SD**	**Availability**
Age at Initial Diagnosis	0.285	59.45 ± 13.51	99.7%	61.68 ± 16.12	100.0%
BMI	0.694	25.79 ± 6.13	90.5%	26.22 ± 7.8	74.0%
Height	0.068	167.7 ± 9.06	91.5%	170.55 ± 9.5	76.0%
Weight	0.313	72.89 ± 19.41	92.8%	76.14 ± 22.1	84.0%

In total, data was retrieved for 371 patients in the TCGA-LIHC cohort, 50 of which had sequencing data for adjacent normal liver tissue. Tables 1 and 2 give descriptive information of the overall TCGA-LIHC cohort, as well as the smaller subset of patients with paired samples. The cohort consisted mostly of white and Asian males. At the time of data collection, the majority of patients also presented with relatively early and mild stages of HCC, as indicated by the various measures of tumour stage and inflammation.

HCC risk factors within the TCGA-LIHC cohort were broad. Firstly, the majority of patients were in the normal to obese ranges of body mass index, however only 5% of patients presented with a history of NAFLD. Approximately 31% of patients had a history of alcohol consumption, and 41% of patients showed evidence of HBV or HCV infection. Thus, the most prominent risk factors in this cohort are viral hepatitis, followed by various obesity and alcohol related risks.

The subset cohort of patients with paired tumour/normal samples taken was somewhat representative of the larger cohort. The smaller cohort did not significantly differ from the larger in terms of gender, age, body mass index, NALFD status, or the pathologic stage of their tumours. However, the smaller cohort included patients with relatively worse prognosis according to Child-Pugh classification, less fibrosis, and were less likely to be affected by viral hepatitis. Additionally, these patients were more likely to be white, and to have a family history of liver cancer.

3.2.2 HORMAD2 Expression

TCGA-LIHC RNASeq data was normalised and analysed for differential expression using the R/Bioconductor package 'DESeq2'. Using paired analysis with 50 matched tumour-normal samples, there was a significant reduction in HORMAD2 expression (DESeq2 $p = 2.11 \times 10^{-6}$) in HCC tumour tissue compared with adjacent normal tissue. Additionally, analysis of the density of HORMAD2 expression showed that tumour tissue displayed mostly low HORMAD2 expression, but also a small cohort with high HORMAD2 expression (**Figure 3.2**).

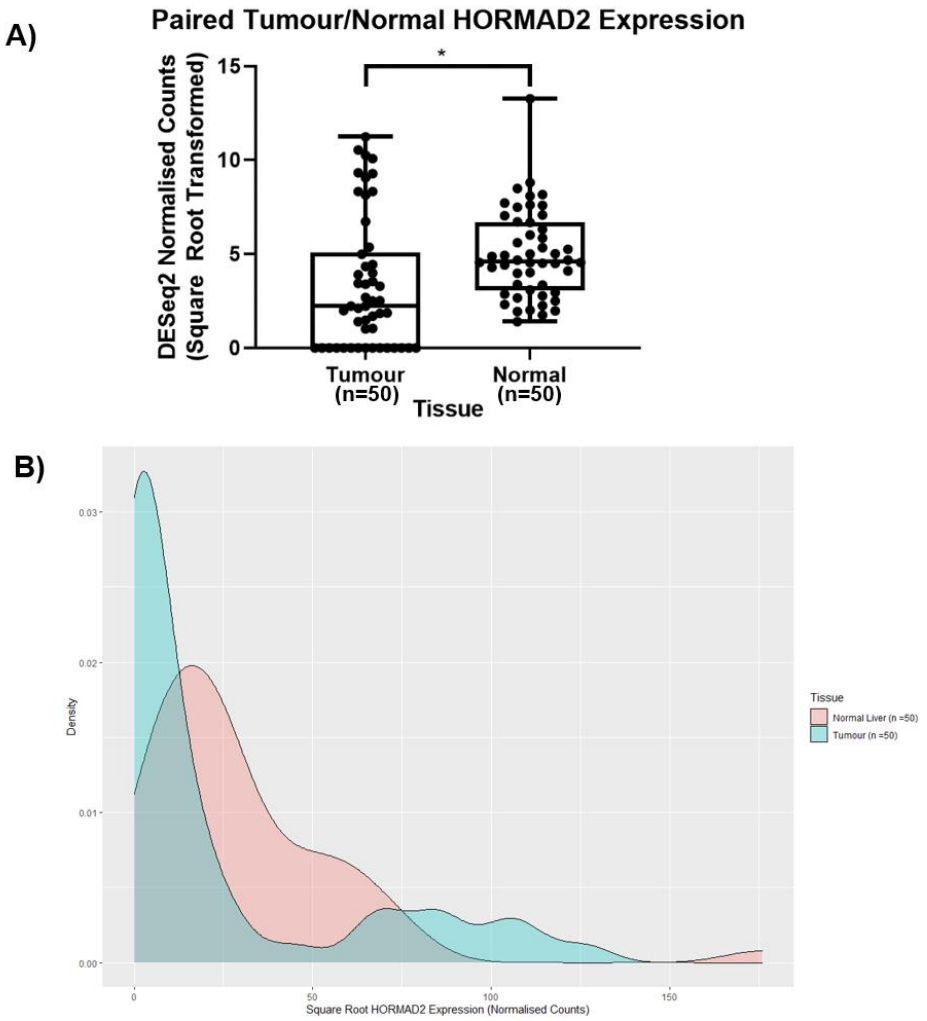

Figure 3.2: HORMAD2 is differentially expressed in tumour and normal tissue. RSEM counts were normalized and analysed for differential expression using DESeq2 paired analysis, including 50 normal samples and 50 tumour samples. **A)** HORMAD2 expression was significantly reduced in tumour vs normal tissue (DESeq2, adjusted $p = 2.11 \times 10^{-6}$). **B)** HORMAD2 expression was plotted as a density plot using ggPlot2. * $p < 0.05$ DESeq2 Wald Test;

Because of the distribution of HORMAD2 expression in patient tumours, the TCGA-LIHC cohort was divided into high, normal and low expression groups for further analysis (hereon referred to as HH, NH, and LH respectively). To determine a cut-off for these groups, a threshold was created at one standard deviation above and below the mean expression in normal liver tissue. The patients with tumours that fell within this threshold for HORMAD2 expression were considered to have normal HORMAD2

expression (NH), while those above and below were considered high (HH) and low (LH) respectively. Using this threshold, the group sizes were: LH = 93, NH = 240, HH = 38.

3.2.3 HORMAD2 Clinical Information

Table 3.3 and **Table 3.4** show the distribution of clinical descriptors across the HORMAD2 expression groups. Select descriptors that indicated a trend with HORMAD2 expression were also plotted against HORMAD2 expression rather than the distinct expression groups (**Figure 3.3**).

Table 3.3: HORMAD2 Categorical Data Table Categorical data separated by HORMAD2 expression groups, low HORMAD2 (LH), normal HORMAD2 (NH) and high HORMAD2 (HH). Percentage (%) values indicate proportion of each expression group falling within each feature type. Significance of distribution of these values was tested using a Chi-Squared test, proportions of each group that were "Not Recorded" were not included in this test. *, $p < 0.05$; **, $p < 0.01$.

		HORMAD2 Expression in Tumour Tissue (Count (%))					
Feature	**Type**	**LH (n = 93)**		**NH (n = 240)**		**HH (n = 38)**	
Adjacent Hepatic Tissue Inflammation	None	27	(29%)	73	(30%)	17	(45%)
Chi-Squared p = 0.273	Mild	27	(29%)	66	(28%)	6	(16%)
	Severe	4	(4%)	11	(5%)	3	(8%)
	Not Recorded	35	(38%)	90	(38%)	12	(32%)
BMI Status	Underweight	7	(8%)	12	(5%)	2	(5%)
Chi-Squared p = 0.242	Normal	45	(48%)	100	(42%)	12	(32%)
	Overweight	19	(20%)	62	(26%)	9	(24%)
	Obese	10	(11%)	47	(20%)	10	(26%)
	Not Recorded	12	(13%)	19	(8%)	5	(13%)
Child Pugh Classification	A	45	(48%)	149	(62%)	23	(61%)
Chi-Squared p = 0.02*	B	5	(5%)	15	(6%)	1	(3%)
	C	0	(0%)	0	(0%)	1	(3%)
	Not Recorded	43	(46%)	76	(32%)	13	(34%)
Embolization Type	Chemoembolization	7	(8%)	15	(6%)	7	(18%)
Chi-Squared p = 0.03*	Radioembolization	2	(2%)	5	(2%)	0	(0%)
	Not Recorded	84	(90%)	220	(92%)	31	(82%)
Fibrosis Ishak Score	0 - No Fibrosis	18	(19%)	44	(18%)	12	(32%)
Chi-Squared p = 0.08	1,2 - Portal Fibrosis	5	(5%)	26	(11%)	0	(0%)
	3,4 - Fibrous Septa	6	(6%)	19	(8%)	3	(8%)
	5 - Nodular Formation and Incomplete Cirrhosis	3	(3%)	6	(3%)	0	(0%)
	6 - Established Cirrhosis	12	(13%)	47	(20%)	11	(29%)
	Not Recorded	49	(53%)	98	(41%)	12	(32%)
Gender	Female	45	(48%)	68	(28%)	8	(21%)
Chi-Squared p = 0.001**	Male	48	(52%)	172	(72%)	30	(79%)
History of Alcohol Consumption	No	66	(71%)	164	(68%)	24	(63%)
Chi-Squared p = 0.681	Yes	27	(29%)	76	(32%)	14	(37%)
History of HBV or HCV	No	63	(68%)	127	(53%)	28	(74%)
Chi-Squared p = 0.007**	Yes	30	(32%)	113	(47%)	10	(26%)
History of NAFLD	No	90	(97%)	227	(95%)	35	(92%)
Chi-Squared p = 0.514	Yes	3	(3%)	13	(5%)	3	(8%)

Pathologic Stage	Stage I	34	(37%)	119	(50%)	18	(47%)
Chi-Squared p = 0.021*	Stage II	18	(19%)	61	(25%)	7	(18%)
	Stage III	32	(34%)	44	(18%)	9	(24%)
	Stage IV	3	(3%)	2	(1%)	0	(0%)
	Not Recorded	6	(6%)	14	(6%)	4	(11%)
Ethnicity	American Indian or Alaska Native	1	(1%)	1	(0%)	0	(0%)
Chi-Squared p = 0.276	Asian	41	(44%)	108	(45%)	9	(24%)
	Black or African American	4	(4%)	11	(5%)	2	(5%)
	White	46	(49%)	112	(47%)	26	(68%)
	Not Recorded	1	(1%)	8	(3%)	1	(3%)
Treated with Radiation Therapy	No	50	(54%)	167	(70%)	23	(61%)
Chi-Squared p = 0.12	Yes	2	(2%)	2	(1%)	0	(0%)
	Not Recorded	41	(44%)	71	(30%)	15	(39%)
Relatives with Family History	No	50	(54%)	140	(58%)	18	(47%)
Chi-Squared p = 0.485	Yes	27	(29%)	71	(30%)	14	(37%)
	Not Recorded	16	(17%)	29	(12%)	6	(16%)
Therapy Type	Ancillary	0	(0%)	1	(0%)	0	(0%)
Chi-Squared p = 0.012*	Chemotherapy	14	(15%)	13	(5%)	4	(11%)
	Other, specify in notes	2	(2%)	0	(0%)	0	(0%)
	Targeted molecular therapy	0	(0%)	5	(2%)	1	(3%)
	Not Recorded	77	(83%)	221	(92%)	33	(87%)

Table 3.4: HORMAD2 Numerical Data Table Numerical data separated by HORMAD2 expression groups, low HORMAD2 (LH), normal HORMAD2 (NH) and high HORMAD2 (HH). Significance of distribution of these values was tested using a Single-Way ANOVA. ***, $p < 0.001$.

		Mean ± SD		
Clinical Data	**p**	**LH (n = 93)**	**NH (n = 240)**	**HH (n = 38)**
Age at Initial Diagnosis (yrs)	**0.0003*** **	55.24 ± 16.19	60.22 ± 12.48	64.84 ± 9.51
BMI (kg/m^2)	0.0527	24.44 ± 5.25	26.11 ± 6.33	27.04 ± 6.56

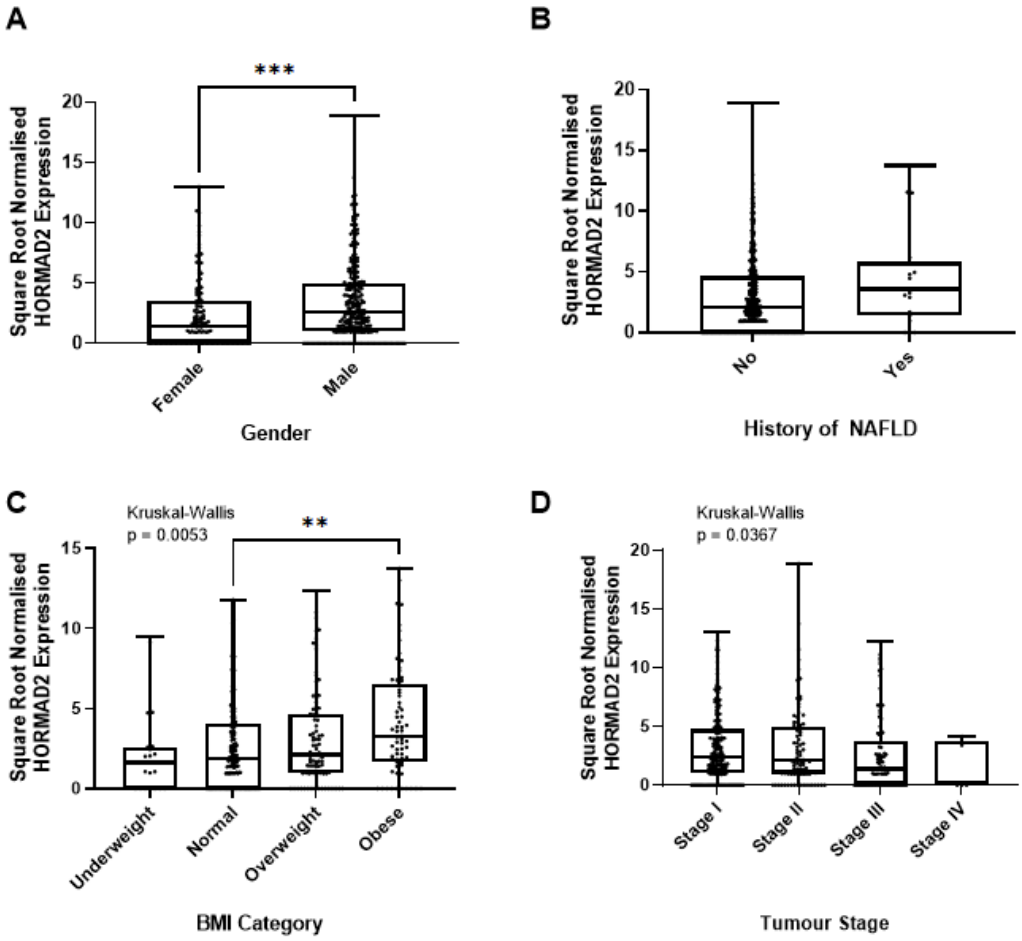

Figure 3.3: Association of HORMAD2 expression with specific clinical features. A) HORMAD2 expression was significantly higher in male patients, **B)** and not significantly affected by history of NAFLD (Mann-Whitney). **C)** HORMAD2 was significantly associated with increasing BMI, and **D)** and tumour stage (Kruskal-Wallis/Dunn's). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Evident in **Table 3.3**, Child-Pugh classification of prognosis was weakly affected by HORMAD2 expression group, however the values in this data were below the statistical power of the Chi Squared test used. Additionally, patients in the high HORMAD2 expression group were less likely to be affected by viral hepatitis. Patients with higher HORMAD2 expression were statistically more likely to be male, which was also supported by the data in **Figure 3.3**. Both **Table 3.3** and **Figure 3.3** indicate that HORMAD2 expression was significantly correlated with tumour stage, where HORMAD2 expression trended towards decreased expression81 at more advanced tumour stages. This change was significant across all groups (Kruskal-Wallis, $p = 0.0367$) but not in any of the multiple comparisons (Dunn's $p > 0.05$). HORMAD2 expression tended to be higher in obese patients (BMI > 30) compared to those with a normal BMI (BMI between 18.5 and 25) (Kruskal-Wallis and Dunn's, $p = 0.0062$). Additionally, patients with higher HORMAD2 expression were significantly older than those with low expression. Finally, HORMAD2 expression was also higher, though non-significantly, in patients with a history of NAFLD (Mann-Whitney, $p = 0.1053$).

3.2.4 Survival

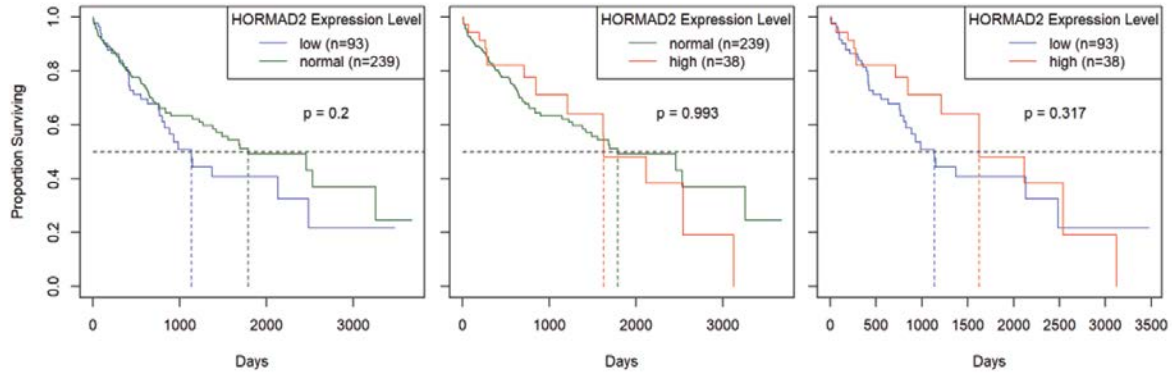

Figure 3.4: HORMAD2 expression is not associated with patient survival. Survival analysis was performed using the R package 'Survival'. Patient cohort was divided into three groups of High (n = 38), Normal (n = 239), and Low (n = 93) HORMAD2 expression. P values on graph are for log-rank test performed in Survival package.

Using patient death and follow-up data, survival analysis was performed to assess whether HORMAD2 expression had an effect of patient survival. Patient survival was not significantly affected by HORMAD2 expression, however LH patients tended to have worse survival rates compared to NH and HH patients.

Given the presence of several potential confounding factors illustrated by **Table 3.3** and **Table 3.4**, a multivariate Cox regression analysis was performed, with the following covariates: HORMAD2 expression group, age, pathologic stage, and gender. This analysis showed that patients with lower HORMAD2 possessed greater risk, however the effect was not significant. Indeed, the only covariate which significantly affected patient survival was pathologic stage.

Table 3.5: Multivariate Cox Regression. Survival data was re-examined using a Cox regression model. Higher risk group was determined based on the nature of the input data and the sign (positive or negative) of the Cox coefficient. 30 patients were eliminated from this test due to missing data (n = 346). ***, $p < 0.001$. CI, confidence interval.

	Coefficient	Higher Risk Group	Hazard Ratio (95% CI)	p
HORMAD2 Expression Group	-0.139	Lower HORMAD2 Expression	0.87 (0.64-1.19)	0.381
Age	0.010	Higher Age	1.01 (1-1.03)	0.169
Pathologic Stage	0.507	Later Tumour Stage	1.66 (1.36-2.03)	8.33×10^{-7} ***
Gender	-0.156	Female	0.86 (0.58-1.26)	0.43
Overall Significance	**p**			
Likelihood ratio Test	1×10^{-5} ***			
Wald Test	7×10^{-6} ***			
Log-Rank Test	4×10^{-6} ***			

3.2.5 Genetic Mutations

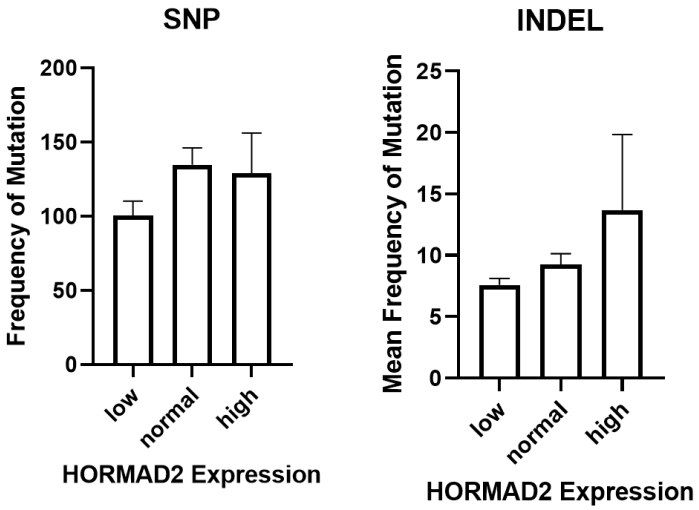

Figure 3.5: Association of HORMAD2 expression with mutation frequency as a measure of genomic stability. INDEL mutations are the sum of insertions and deletions.

Given HORMAD2's known interactions with several DNA-repair proteins, it was hypothesised that HORMAD2 expression could affect genomic stability. As a measure of genomic stability, the overall level of SNPs and insertion/deletions (indels) were each measured for each HORMAD2 expression group (**Figure 3.5**). There was no significant effect observed between the HORMAD2 expression group and the level of genomic stability ($p > 0.05$).

Next, mutated genes were investigated more specifically. Only one patient in the TCGA-LIHC cohort had a mutant form of HORMAD2, and this mutation was in a non-coding region of the gene. Thus, based on this analysis it is unlikely that HORMAD2 mutations play a significant role in liver cancer.

However, the mutation data can still be analysed to elucidate patterns of replaced function or regulation by other proteins. To investigate this, mutation data was combined with HORMAD2 RNASeq expression data to investigate whether mutations in particular pathways was associated with dysregulated HORMAD2 expression. Mutation data for each gene per patient was retrieved and filtered to remove all silent mutations. The Fisher's Exact test was used to determine significant differences in mutation frequency between the HORMAD2 expression groups of patients. Genes with significantly enhanced mutation frequencies in either high or low HORMAD2 tumours ($p < 0.05$) were retrieved **(Appendix A and Appendix B)**. These lists were analysed for pathway enrichment (**Figure 3.6**).

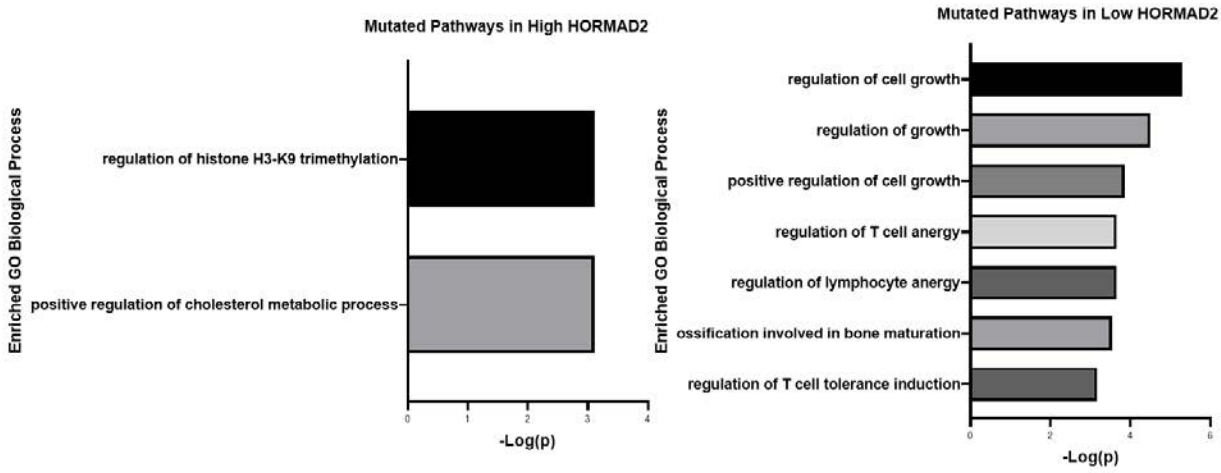

Figure 3.6: Genetic pathways with enriched mutations. Mutation frequencies per gene were calculated for patients with high or low HORMAD2 expression and compared with the normal HORMAD2 expression group using Fisher's Exact Test. Genes with significantly higher mutation rates in each group were parsed through GORILLA for gene ontology enrichment analysis.

In HH patients, mutations were enriched in genes involved in histone H3K9 trimethlyation, and cholesterol metabolism. This enrichment was due to mutations in alpha thalassemia/mental retardation syndrome x-linked (ATRX) and sirtuin 1 (SIRT1), as well as SREBF chaperone (SCAP) and carboxylesterase 1 (CES1) respectively. However, other proteins of interest were also present in the list of highly mutated genes **(Appendix A)**. These included apoptosis and cell cycle regulation genes (ACVR2A, STAG1, CASP3, NFKB1, BIRC6, ACIN1, E2F3), histone methylation regulators (DOT1L, ATRX), and metabolic proteins (SIRT1, G6PC).

Conversely, LH patients harboured mutations in genes involved in cell growth. Specific genes of interest included proteins involved in DNA repair (BAP1) cytokine signalling (SOCS6), cell cycle regulation (MAP4K5, KMT2E), and deubiquitination (USP19) **(Appendix B)**.

3.2.6 Co-Expression

Another way to elucidate protein function from RNA expression data is to investigate genes and genetic pathways which positively or negatively correlate with HORMAD2 expression. To evaluate this, a Spearman correlation was performed for HORMAD2 with every other gene in the dataset. These genes were then ranked by the strength of their positive or negative correlation (Spearman Rho value)**(Appendix C)** and these ranked lists were used to identify enriched genetic pathways (**Figure 3.7**).

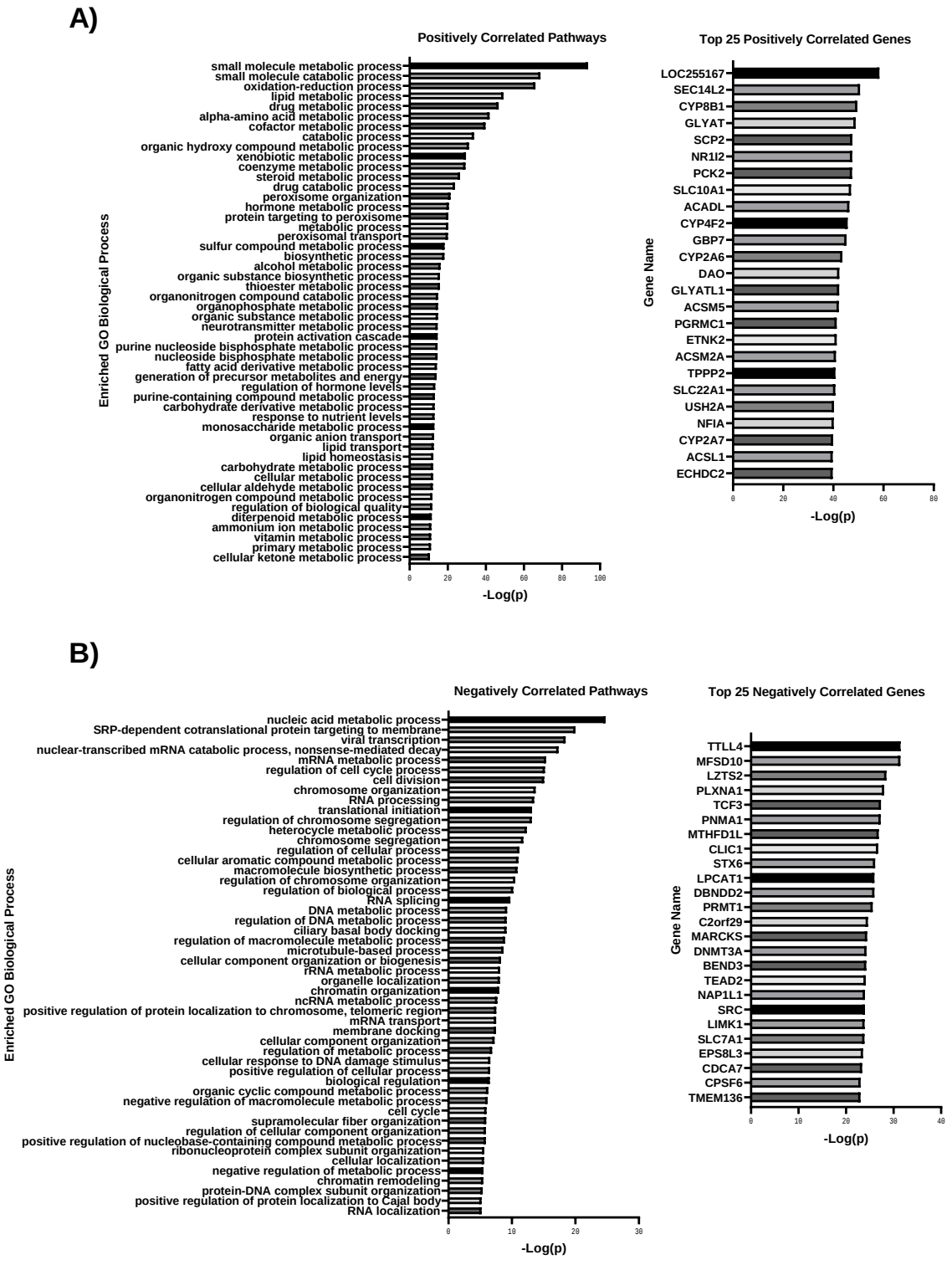

Figure 3.7: Genetic pathways co-expressed with HORMAD2. Co-expression was evaluated using Spearman correlation of RNASeq expression counts. After ranking genes by their Rho value, pathway enrichment was performed using GORILLA. This analysis was performed with in order of both the **A)** strongest positive correlation, and **B)** the strongest negative correlation. Note: p-values for individually correlated genes are not corrected for multiple testing as they are for the purpose of ranking.

HORMAD2 expression positively correlated with genes involved in the metabolism of lipids and cholesterol, amino acids, carbohydrates, and alcohol. Additionally, genes involved in oxidative-reduction reactions were prevalent. Of the top 25 positively correlated genes, four members of the cytochrome P450 complex were present, as well as other proteins which interact with the P450 complex such as PGRMC1. Furthermore, a large number of genes were involved in responses to various forms of cellular stress such as oxidative stress, xenobiotics, bacteria, and toxic small molecules. The gene with the strongest correlation to HORMAD2 expression was a non-protein coding RNA, LOC255167.

HORMAD2 expression was negatively correlated with genes involved in the cell cycle and cell growth. The most highly correlated genes were less defined in terms of function compared to the positively correlated genes, however genes of interest included anti-apoptotic genes like MFSD10, and cell-cycle related genes such as TTLL4, LZTS2, MARCKS, NAPL1L1, SRC, and CDCA7.

3.2.7 DNA Methylation

Given that HORMAD2 expression is tissue-restricted, epigenetic DNA methylation is hypothesised to play a role in its regulation. DNA methylation occurs specifically at sites in the genome where a cytosine is followed by a guanine nucleotide (CpG) and regions of the genome with high frequency of CpG sites are called CpG islands. CpG islands are often found at the beginning of genes and may be important in regulating gene expression. UCSC data identified 16 CpG sites for the HORMAD2 gene, a number of which are within a CpG island in the supposed promoter of HORMAD2 (**Table 3.6**).

Table 3.6: CpG sites investigated for affects on HORMAD2 expression. This list was retrieved from UCSC. Positions are for Hg19

CpG ID	Position	Location Description
cg06623197	chr22:30400762	Upstream
cg15797101	chr22:30422326	Upstream
cg01000280	chr22:30472635	Upstream (Promoter)
cg00334274	chr22:30474914	Upstream (Promoter)
cg21890667	chr22:30476088	5' CpG Island
cg16686158	chr22:30476097	5' CpG Island
cg04046669	chr22:30476205	5' CpG Island
cg15209808	chr22:30476253	5' CpG Island
cg14509403	chr22:30476280	5' CpG Island
cg21843594	chr22:30476284	5' CpG Island
cg17632937	chr22:30476325	5' CpG Island
cg23268208	chr22:30476344	5' CpG Island
cg01141459	chr22:30476451	5' CpG Island
cg13245431	chr22:30476524	5' CpG Island
cg10230314	chr22:30552800	Exon
cg24211826	chr22:30572325	Exon

Three different approaches were taken to evaluate whether each CpG had a significant effect on HORMAD2 expression. First, the methylation level at each CpG was compared between paired samples of tumour tissue and adjacent normal liver tissue (**Figure 3.8**). Because HORMAD2 is down-regulated in tumours, methylation at important CpGs would be expected to be higher in tumour tissue.

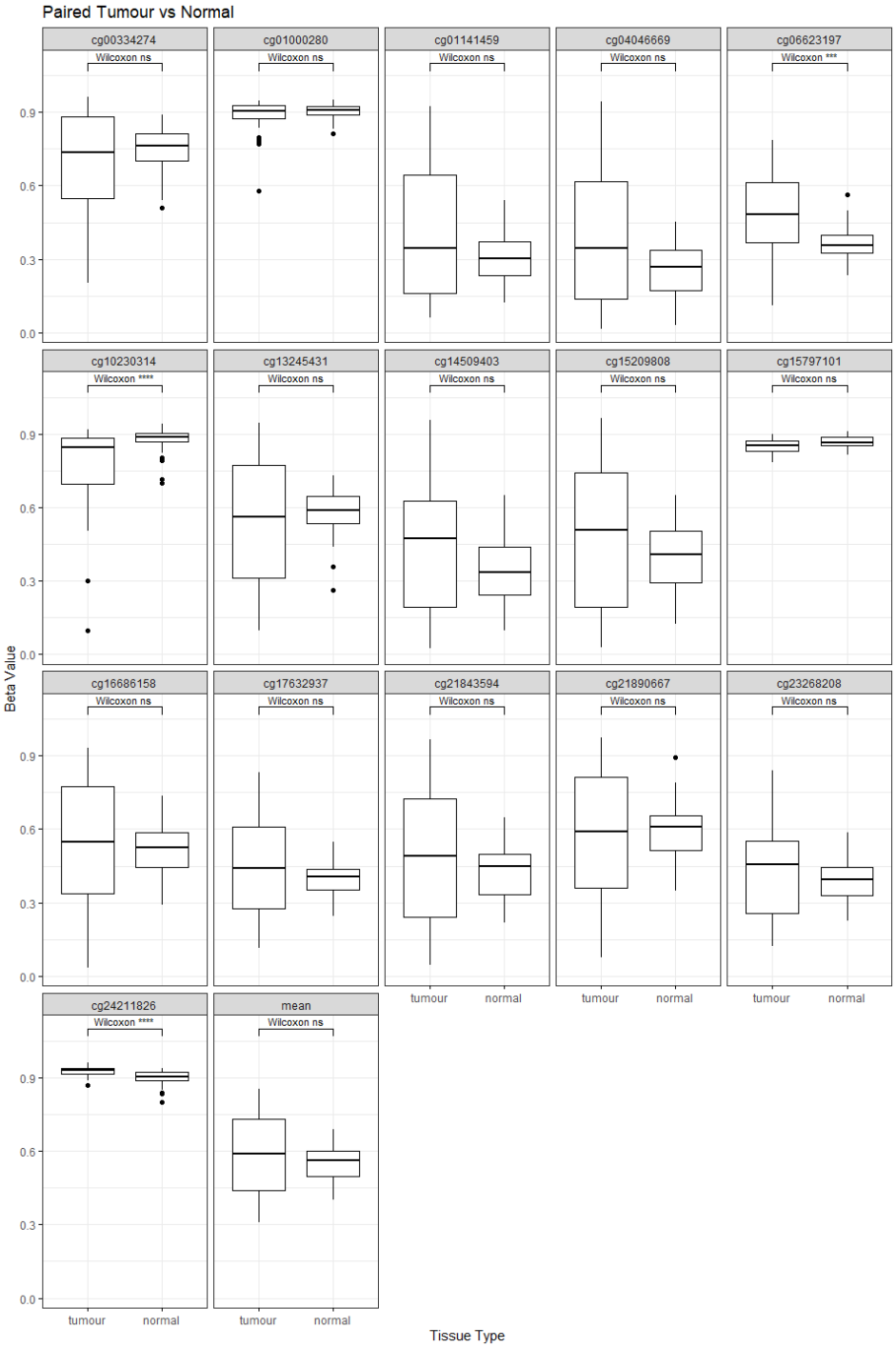

Figure 3.8: Tumour vs normal methylation level at CpGs of interest. Methylation level is in the form of beta value, with 0 being unmethylated and 1 being fully methylated. Significance values calculated using Wilcoxon test with Bonferroni correction for multiple testing. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Indeed, a number of CpG sites were slightly up-regulated in tumour tissue, however most of these differences were insignificant.

Next, the methylation at each CpG was compared in the tumour tissue between LH, NH and HH patients (**Figure 3.9**). As expected, many CpG sites were hypermethylated in LH patients, and hypomethylated in HH patients.

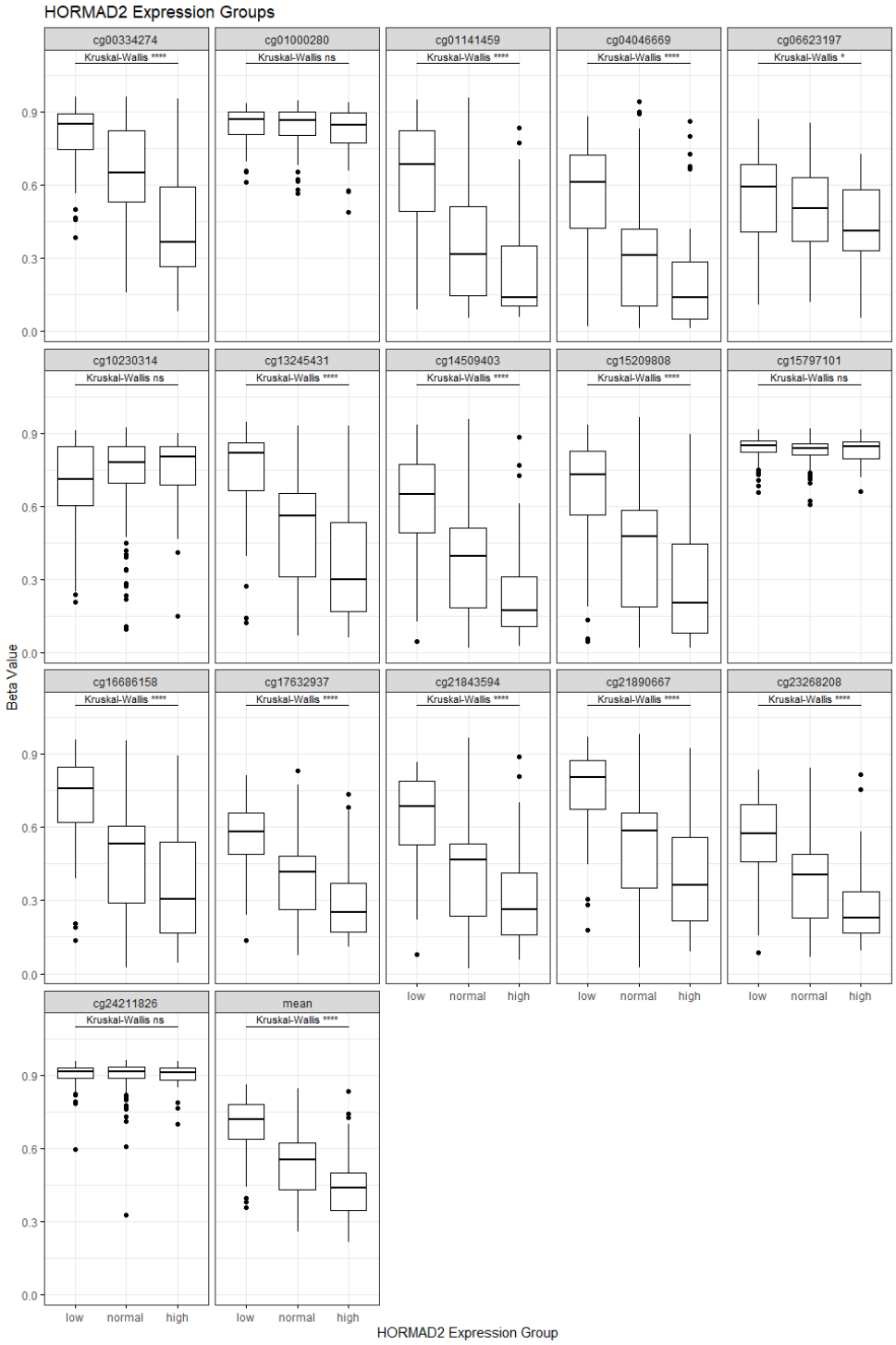

Figure 3.9: Association of HORMAD2 Expression with methylation level at CpGs of interest. Methylation levels were plotted for patients in the low, normal and high HORMAD2 groups. Methylation level is in the form of beta value, with 0 being unmethylated and 1 being fully methylated. Significance values calculated using Kruskal-Wallis test with Bonferroni correction for multiple testing. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Finally, the methylation level at CpG site was compared with the HORMAD2 RNA expression level of each patient and a Pearson correlation analysis was performed. Both tumour and normal samples were included (**Figure 3.10**).

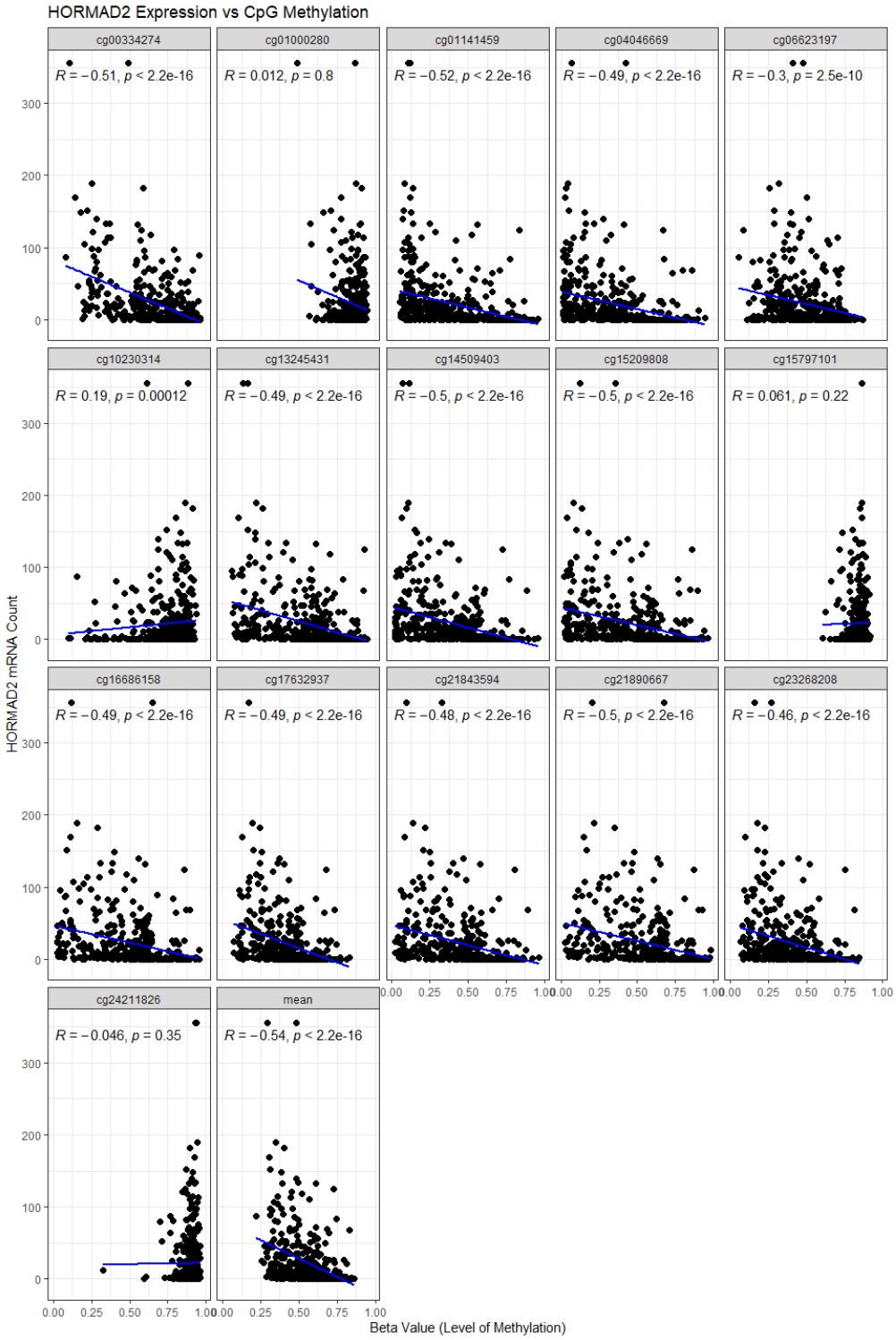

Figure 3.10: Correlation of HORMAD2 Expression with methylation level at CpGs of interest. Methylation level for each sample was plotted against individual HORMAD2 expression. Correlation analysis was performed using Spearman correlation. Methylation level is in the form of beta value, with 0 being unmethylated and 1 being fully methylated.

Based on the three analyses above, eleven CpG sites that appeared to have the strongest effect on HORMAD2 expression were manually selected to create a CpG profile for HORMAD2 regulation (**Table 3.7**). Ten of these CpG sites were within the 5' CpG island.

Table 3.7: CpG profile for HORMAD2 regulation. CpG sites which had the strongest effect on HORMAD2 expression were manually selected.

CpG ID	Position	Location Description
cg06623197	chr22:30400762	Upstream
cg21890667	chr22:30476088	5' CpG Island
cg16686158	chr22:30476097	5' CpG Island
cg04046669	chr22:30476205	5' CpG Island
cg15209808	chr22:30476253	5' CpG Island
cg14509403	chr22:30476280	5' CpG Island
cg21843594	chr22:30476284	5' CpG Island
cg17632937	chr22:30476325	5' CpG Island
cg23268208	chr22:30476344	5' CpG Island
cg01141459	chr22:30476451	5' CpG Island
cg13245431	chr22:30476524	5' CpG Island

The three analyses were then performed again on the mean methylation level across all of the CpG sites in the CpG profile (**Figure 3.11**). There was no significant difference in CpG methylation between tumour and normal samples. However, CpG sites within the HORMAD2 profile were significantly hypermethylated in LH patients, and hypomethylated in HH patients. Additionally, there was a significant inverse correlation between HORMAD2 expression and extent of methylation. In sum, these findings suggest that HORMAD2 expression is at least partly regulated by CpG methylation in human HCC.

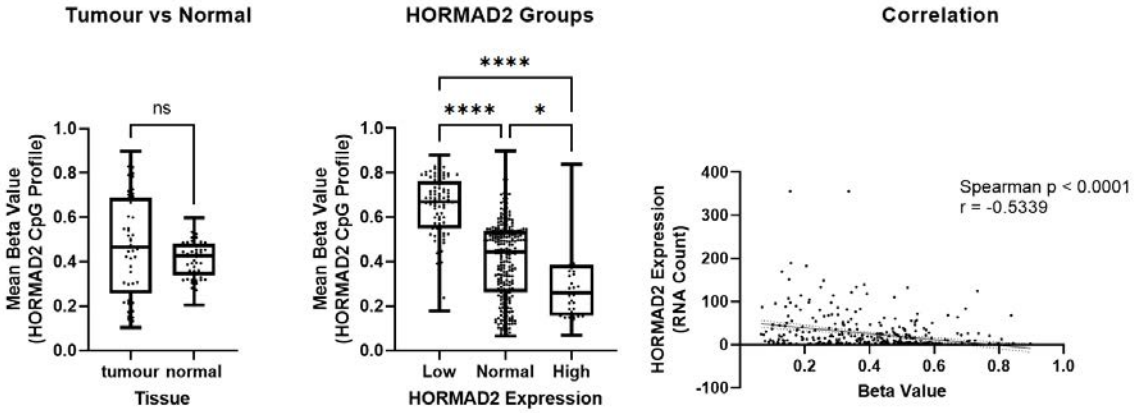

Figure 3.11: Regulation of HORMAD2 expression by methylation. Methylation values are based on the average of beta values (0 being unmethylated and 1 being fully methylated) across the CpG profile generated for HORMAD2 expression. Data is shown for: **left**) methylation in tumour and normal tissue (Wilcoxon), **middle**) methylation in each cohort of HORMAD2 expression (Kruskal-Wallis), **right**) correlation analysis (Spearman) between HORMAD2 expression and CpG methylation. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

3.2.8 miRNA

Another common form of gene regulation important for HCC is microRNA (miRNA) activity. To investigate potential miRNA regulation of HORMAD2, miRNA expression data for TCGA-LIHC was retrieved from the GDAC firehose pipeline. The miRNA's with a significant negative correlation (Spearman rho < 0, $p < 0.05$) with HORMAD2 expression were selected, and these were further filtered to only miRNAs with either validated or predicted interactions with HORMAD2 based on a multiMiR search. Those with significant correlation ($p < 0.05$) are listed in **Figure 3.12**.

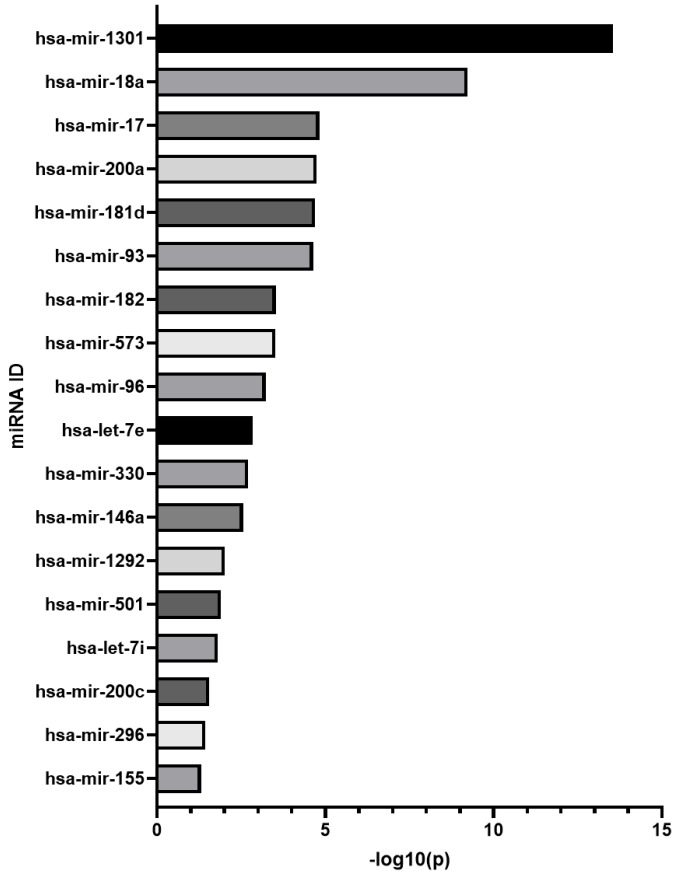

Figure 3.12: miRNAs correlated with HORMAD2 expression. Significant miRNAs were those with the strongest negative correlation (Spearman Rho) with HORMAD2 expression across 424 samples in the TCGA-LIHC data (tumour and normal combined) and were further filtered to those with predicted activity against HORMAD2 using MultiMiR. –log10(p) was generated from Spearman correlation p value and corrected using the Bonferroni method.

In total, 18 miRNAs were identified with both significant HORMAD2 negative correlation, and either predicted or validated targeting to HORMAD2. Two of the negatively correlated genes were listed in either the tarbase or mirtarbase databases as "Validated" for targeting HORMAD2, and these were both detected by the weakest class of evidence including microarray and next-generation sequencing. These miRNAs were hsa-mir-182 ($r = -0.25$, $p = 2.94 \times 10^{-4}$), hsa-mir-146a ($r = -0.22$, $p = 0.002750318$). In addition to these, some unvalidated but predicted miRNAs had particularly strong correlations, including hsa-mir-1301 ($r = -0.40$, $p = 2.78 \times 10^{-14}$) and hsa-mir-18a ($r = -0.34$, $p = 5.86 \times 10^{-10}$).

3.2.9 Transcription Factor Prediction

The 2kb region upstream of the HORMAD2 transcription start site was analysed for predicted transcription factor binding sites using PROMO v3 (200, 201). Those binding sites with <5% dissimilarity between their true sequence and the sequence of the genome were retrieved and are listed in **Table 3.8**. Predicted transcription factors of interest were involved in the DNA damage response (p53, YY1, c-Ets-1, c-jun, E2F-1), inflammation and/or stress response factors (NFκB1,), liver metabolism (PPARα, HNF-1A, GR), and the oestrogen receptor (ERα).

Table 3.8: Predicted transcription factor binding sites in HORMAD2 promoter. Prediction was performed by PROMO v3, filtering for <5% dissimilarity.

Factor name	Dissimilarity	RE equally	RE query
GCF [T00320]	0	0.0916	0.16552
NF-1 [T00539]	0	0.12213	0.15361
c-Ets-1 [T00112]	0	0.24426	0.25838
HNF-1A [T00368]	0	0.24426	0.14176
GR [T05076]	0	0.36639	0.25915
p53 [T00671]	0	0.36639	0.48731
IRF-2 [T01491]	0	0.48853	0.40091
STAT4 [T01577]	0	0.48853	0.40091
AP-2alphaA [T00035]	0	0.97705	1.33892
XBP-1 [T00902]	0	0.97705	0.87517
Pax-5 [T00070]	0	1.09918	1.6514
TFIID [T00820]	0	1.09918	0.67268
FOXP3 [T04280]	0	1.46558	1.3989
TFII-I [T00824]	0	1.46558	1.51556
ER-alpha [T00261]	0	1.9541	2.07904
GR-beta [T01920]	0	3.9082	2.72545
GR-alpha [T00337]	0	7.81641	7.68136
YY1 [T00915]	0	7.81641	7.68448
C/EBPbeta [T00581]	0	15.63281	14.34263
Elk-1 [T00250]	0.134348	0.06107	0.06986
GATA-1 [T00306]	0.280028	0.97705	0.6789
Sp1 [T00759]	0.574521	0.00763	0.01612
RXR-alpha [T01345]	0.848226	0.48853	0.56853
POU2F1 [T00641]	0.929531	0.01384	0.00956
HIF-1 [T01609]	1.005355	0.09923	0.12086
T3R-beta1 [T00851]	1.110682	0.07633	0.0836
c-Ets-2 [T00113]	1.64415	0.0458	0.04553
Ik-1 [T02702]	2.374299	0.00063	0.00082
C/EBPalpha [T00105]	2.441016	0.48853	0.43852
ETF [T00270]	2.623453	0.00716	0.01567
NF-AT1 [T01948]	2.756277	0.05916	0.04688
PR A [T01661]	2.80933	0.73279	0.52911
PR B [T00696]	2.80933	0.73279	0.52911
STAT1beta [T01573]	2.898434	0.05152	0.0445
c-Jun [T00133]	3.049104	0.24426	0.23855

TCF-4E [T02878]	3.151193	0.24426	0.20188
MAZ [T00490]	3.175881	0.00474	0.00743
ATF3 [T01313]	3.372402	0.0916	0.08004
NF-kappaB1 [T00593]	3.384737	0.01813	0.02721
HNF-3alpha [T02512]	3.500065	0.2748	0.1533
CREB [T00163]	3.500888	0.10686	0.10922
NFI/CTF [T00094]	3.793671	0.1832	0.20621
PPAR-alpha:RXR-alpha [T05221]	3.872523	0.02576	0.03302
NF-Y [T00150]	4.186615	0.1832	0.15555
HOXD10 [T01425]	4.321431	0.03435	0.01767
HOXD9 [T01424]	4.321431	0.03435	0.01767
USF2 [T00878]	4.528187	0.0687	0.07958
E2F-1 [T01542]	4.545253	0.15266	0.20389
c-Myb [T00137]	4.974489	0.24426	0.2222

Additionally, **Table 3.9** lists the binding sites of predicted transcription factors which align either nearby or exactly with significantly methylated CpG sites from **Table 3.7**.

Table 3.9: Transcription factor sites in close proximity to methylated CpGs. The 'Exact Match' column indicates whether methylated CpG sites fell within the transcription factor binding site, as opposed to being within close proximity. TF; Transcription factor.

CpG Site	CpG Position	Transcription Factor	TF-Binding Site	Exact Match?
cg21890667	chr22:30476088	p53	chr22:30476083-30476089	Yes
cg16686158	chr22:30476097	RXR-alpha	chr22:30476107-30476113	No
cg04046669	chr22:30476205	ER-alpha	chr22:30476193-30476197	No
		RXR-alpha	chr22:30476210-30476216	No
cg15209808	chr22:30476253	Pax-5	chr22:30476250-30476256	Yes
		p53	chr22:30476250-30476256	Yes
cg14509403	chr22:30476280	p53	chr22:30476266-30476272	No
cg21843594	chr22:30476284	TCF-4E	chr22:30476293-30476299	No
cg17632937	chr22:30476325	RXR-alpha	chr22:30476346-30476352	No
cg23268208	chr22:30476344	ER-alpha	chr22:30476347-30476351	No
cg01141459	chr22:30476451	ER-alpha	chr22:30476445-30476449	No
cg13245431	chr22:30476524	Pax-5	chr22:30476518-30476524	Yes
		p53	chr22:30476518-30476524	Yes

Significant sites of HORMAD2 CpG methylation were present exactly within binding sites for p53 and Pax-5 at 3 different sites. Additionally, methylation sites were in very close proximity to binding sites for RXR-alpha, ER-alpha and TCF-4E.

3.2.10 Molecular Interactions

Numerous databases exist online which document the results of high-throughout protein-protein interaction studies. These include Interactome3D (202), Targetmine (203), The BioGrid (204). Searching these databases, HORMAD2 was found to have experimental evidence for interactions with a number of other proteins, including: NSD2/WHSC1 (Interactome, Targetmine), SYCP2 (Targetmine), and ESR2 (TheBioGrid).

3.2.11 Protein Features

A number of online tools were also used to predict various features within the HORMAD2 sequence. For example, the Rostlab NLSdb (205) and the NLS mapper (206) and ELM (207) tools all detected a nuclear localisation signal at the C-terminal end of the HORMAD2 sequence. Additionally, the ELM tool detected a MAPK-interacting sequence in the C-terminus, two S/T-Q sites for phosphorylation by PIKK family members, a BRCT-domain interaction peptide, and three sites for interactions with the de-ubiquitination enzyme USP7. ProTStab (208) was used to predict a melting point for HORMAD2 of 66.654°C. Finally, the PredictProtein database (209) predicted a disordered C-terminal tail in the HORMAD2 sequence, as well as a DNA-binding site at the N-terminus.

3.2.12 Homology-Based Structural Modeling

Based on SAVES analysis, SWISS-MODEL (214) was found to produce the best predicted 3-Dimensional structure for HORMAD2. However, this model was based on a template with only 33% sequence similarity and only 76% coverage of the protein was achieved. This produced a protein structure with relatively poor quality indicators (GMQE, QMEAN). This structure is shown below (**Figure 3.13**). This protein structure was parsed through ProFunc (215), a bioinformatic tool which predicts protein function based on 3-dimensional structure. Unfortunately, this produced no significant predictions.

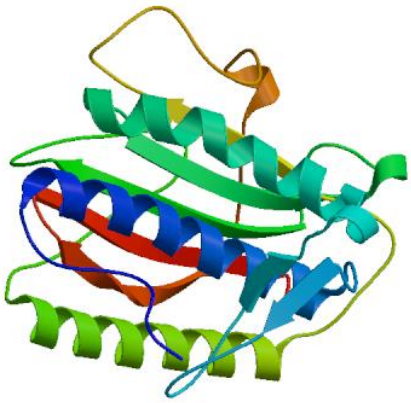

Figure 3.13: Predicted HORMAD2 structure. Modelled by SWISS-MODEL based on homology with *C. elegans* HIM-3 and HTP-3 (4tzj.1.A). GMQE = 0.48, QMEAN = -2.03

3.3 Discussion

The majority of patients in the TCGA-LIHC cohort were Caucasian, in the range of normal to obese BMI, and had primary HCC risk factors of alcohol consumption and viral hepatitis. Very few of the patients had a history of NAFLD. As the preliminary data leading to this investigation suggests a role of HORMAD2 in NAFLD-related HCC, this means that the TCGA-LIHC cohort does not represent an ideal study group for the characterisation of HORMAD2. Additionally, there are several significant differences in the subset patients from which paired normal samples were taken when compared with the entire cohort, including significant differences in prognosis, fibrosis, viral hepatitis status, ethnicity, and family history of cancer (**Table 3.1**). This is likely to cause confounding issues in the data. Nonetheless, the size of the cohort and the extent of information available still makes it an invaluable resource to guide further investigation into HORMAD2 function.

HORMAD2 is expressed in the human liver, and is down-regulated in HCC

HORMAD2 was found to be expressed in almost all patient tumours, as well as in surrounding healthy liver tissue. This supports previous findings of HORMAD2 expression in liver tissue (40). Furthermore, HORMAD2 was found to be significantly down-regulated in most HCC tumours. In actuality, this down-regulation was due to a vast number of tumour samples without any HORMAD2 expression at all, which was not observed at all in the healthy liver samples (**Figure 3.2**). This is consistent with previous findings that HORMAD2 expression decreases with advancing NAFLD severity, potentially implicating it in the mechanism of NAFLD-induced HCC (50). Additionally, these findings reflect those of Lin (2018) who found that HORMAD2 was hypermethylated and down-regulated in thyroid cancer (44). However, these results are also in direct contrast with studies of human lung cancer, which show aberrant HORMAD2 expression (40, 48).

HORMAD2 is classified as a cancer/testis antigen (40), which suggests that its expression is usually limited to the testes and aberrant expression occurs in tumourigenesis. While this pattern of expression appears to be the case in lung cancer, for which this classification was proposed, the results presented here do not indicate that HORMAD2 is a cancer/testis antigen. Instead, these results suggest that HORMAD2 plays a functional role in the liver, and its expression is lost during tumourigenesis. This may be as a result of changes in gene regulation, or alternatively the loss of HORMAD2 expression itself could produce a cancer-promoting environment.

Finally, it should be noted that, while HORMAD2 expression was decreased on average in HCC tissue, there were also a number of patients with high expression in their tumour tissue. For this reason, the subjects were separated into three groups of low, normal, and high HORMAD2 expression for further analyses.

HORMAD2 expression is higher in male tumours

A number of descriptive statistics were available for each patient and were analysed for correlations with HORMAD2 expression. HORMAD2 expression was significantly higher in male tumours than female tumours. This is of interest for a number of reasons. For example, HORMAD2 is also expressed much more dominantly in the testes than in ovaries, and knockout of HORMAD2 in the ovaries appears to have no effect (33). As discussed previously, this may be a result of the differing consequences of MSUC and MSCI, however the question still remains as to why HORMAD2 expression is lower in the ovaries. Secondly, males have been proven to be at far greater risk of HCC than females, and this increased risk is likely to be due to both behavioural and biological differences (216). Thus, any sources of gender disparate molecular mechanisms of HCC are of significant interest. However, given that the overall difference in expression was small, and the fact that other evidence suggests a tumour-suppressive role for HORMAD2 rather than oncogenic, this difference is likely to be inconsequential.

HORMAD2 expression correlates with tumour stage and BMI, but not patient survival

HORMAD2 expression was higher in patients with higher BMI. This is of interest because patients with high BMI are more likely to proceed into pro-inflammatory, NAFLD-inducing states, and this suggests that HORMAD2 expression could be in part regulated by these factors. Additionally, HORMAD2 expression was found to be significantly reduced in patients experiencing later stages of liver cancer. Again, these results suggest that HORMAD2's function and regulation may be in response to the inflammatory insult associated with obesity, likely in the form of DNA damage given the role of HORMAD2 in meiocytes. Furthermore, given that HORMAD2's expression is decreased in tumours and in later stages of cancer, its function in the liver appears to be anti-tumourigenic.

HORMAD2 expression did not significantly impact patient survival in this cohort. However, patients with lower HORMAD2 expression tended to have decreased survival. This lack of significance could be due to the fact that HORMAD2 on its own does not have a strong enough effect to cause a statistical effect. However, to investigate this further a proportional hazards analysis was performed, additionally considering age, tumour stage and gender in the survival analysis. This analysis revealed again that HORMAD2 expression was not a significant predictor of patient survival, and instead that only tumour stage was a significant predictor. Thus, it is likely that the fact that HORMAD2 expression is lower in more advanced tumours in the TCGA-LIHC cohort has confounded the results of the survival analysis.

HORMAD2 correlated pathways

To further elucidate gene function, HORMAD2 expression was investigated for correlation with genetic pathways in two ways: firstly based the prevalence of mutations in genetic pathways, and then secondly based on the expression of genes in genetic pathways.

HORMAD2 expression was also screened for correlation with mutations in specific genes. It was hypothesised that if other genes were responsible for regulating HORMAD2, then their mutation may cause up or down-regulation of HORMAD2 depending on the nature of this regulatory relationship. Additionally, it was hypothesised that if HORMAD2 shared a similar function to another gene, then a mutation in that gene may cause up-regulation of HORMAD2 for the purpose of replacement of function. However, it should be noted that such mutations could be in the form of gain-of-function or loss-of-function mutations. Thus, this analysis cannot provide a definitive direction of association.

High HORMAD2 expression was most associated with mutations in histone H3K9-trimethylating genes, and in the regulation of cholesterol metabolism. Mutated H3K9 methylating proteins included ATRX and SIRT1. H3K9 methylation is associated with gene silencing. However, besides the obvious conclusion that HORMAD2 may be repressed by histone methylation, this relationship could additionally or alternatively be due to the DNA-repair functions of H3K9 methylation. H3K9 methylation is known to accumulate at sites of DNA DSBs, and it plays an important role in the recruitment of DNA-repair proteins, including the recruitment of ATM via TIP60, and blocking of 53BP1-mediated NHEJ (217). Thus, mutations in this pathway could result in DSB persistence, which may also be a cause for increased HORMAD2 expression.

Additionally, patients with high HORMAD2 expression also had enriched mutations in genes responsible for cholesterol metabolism, as well as gluconeogenesis in the form of G6PC. This is of particular importance because cholesterol accumulation in the liver has been shown to be an important contributor to NAFLD, through activation of Kupffer cells and stellate cells to promote inflammation and fibrogenesis (218). Furthermore, glycogen storage disease, associated with glucose-6-phosphatase defects, is another major cause of NAFLD (219). These further suggest that factors which contribute to NAFLD-progression also cause increased HORMAD2 expression. Finally, high HORMAD2 expression was also associated with mutations in a number of cell-cycle regulators such as CASP3, E2F3, and NFKB1, which could indicate either replacement of function or dysregulation by these genes. In either case, this is indicative of a cell-cycle related function to HORMAD2.

In terms of gene expression correlation, HORMAD2 was positively associated with genes involved in the metabolism of lipids and cholesterol, oxidative-reduction reactions, and in the response to toxins and xenobiotics. The most prominent of these were numerous members of the cytochrome P450 complex. The cytochrome P450 complex is one of two endogenous sources of reactive oxygen species in the liver and a key cancer initiator (67).

Conversely, low HORMAD2 expression was associated mostly with mutations in cell cycle regulators, most notably: SOCS6 which promotes cell death (220); BAP1, a BRCA1-associated de-ubiquitinating protein which promotes cell death in response to DNA damage (221); USP19, another deubiquitinating enzymes which promotes cell growth (222); MAP4K5, a component of the MAP kinase pathway which

may act as a tumour suppressor (223). It is noteworthy that the presence of multiple de-ubiquitinating enzymes in this group suggests that ubiqutination may be involved in HORMAD2 regulation.

In the correlation analyses, an inverse correlation was present between HORMAD2 and pathways required for the cell cycle to proceed, such as cell cycle regulators themselves as well as genes involved in nucleic acid synthesis. This inverse correlation could indicate that HORMAD2 either directly or indirectly down-regulates these processes, or conversely that cell-proliferation-promoting factors down-regulate HORMAD2.

While the results of these analyses are difficult to interpret because of a lack of directionality, hypotheses can be drawn from their trends. Consistently, HORMAD2 expression appears to be positively associated with cellular stress pathways, including oxidative stress from deficiencies in lipid and cholesterol metabolism, DNA damage, and xenobiotics. This could suggest that HORMAD2 is up-regulated simultaneously with these pathways and, in extension, during pro-inflammatory and pro-tumourigenic cellular stress. Additionally, HORMAD2 expression is consistently associated with mutations in cell cycle regulators, with which its expression is negatively correlated. This provides some evidence that HORMAD2 is involved in down-regulation of the cell cycle, and that it may be capable of replacing the function of other cell-cycle regulators when they are lost. Thus, there is evidence to support a role for HORMAD2 in down-regulation of the cell cycle or pro-apoptosis in response to cellular stress, probably in the form of DNA damage.

HORMAD2 regulation

Based on tumour/normal expression, association with HORMAD2 expression groups, and direct correlation with HORMAD2 expression, a profile of key CpG sites was created which significantly inversely correlated with HORMAD2 expression (**Table 3.7**). This suggests that DNA methylation plays a significant role in HORMAD2 regulation. This is consistent with a previous study, which showed that HORMAD2 is significantly hypermethylated in thyroid cancer, and found 9 of 11 of the same CpG sites to be significant factors in this regulation (44).

Additionally, microRNA expression was analysed for negative correlation with HORMAD2 expression, suggestive of regulation. **Figure 3.12** shows a list of microRNAs predicted to regulate HORMAD2 based on negative expression correlation, and also predicted binding sites within the HORMAD2 transcript. The two microRNAs with the strongest negative correlation were hsa-mir-1301 and hsa-mir-18a. Additionally, two of these microRNAs have been previously identified as HORMAD2-interactive in the tarbase or mirtarbase miRNA databases through high throughput screening: hsa-mir-182, hsa-mir-146a.

Interestingly, a previous study found that LOC255167 (aka LINC01018), a long noncoding RNA, acts as a tumour suppressor in HCC by quenching the activity of hsa-mir-182 (224). LOC255167 was also

the gene with strongest positive correlation with HORMAD2 expression of over 16000 genes analysed, with a Spearman Rho value of 0.683 and a p value of $0.467x10^{-59}$ (**Figure 3.7**). Hsa-mir-182 has been previously implicated in HCC, with its expression associated with increased invasion and proliferation (225). Thus, this suggests that hsa-mir-182 could play an important role in HORMAD2 expression and supports role for HORMAD2 in HCC.

Lastly, the upstream promoter region of the HORMAD2 genes was analysed for transcription factor binding sites. Because these binding sites are relatively small, such bioinformatic searches often provide extensive results which are unlikely to all be real. However, some binding sites of interest were statistically more likely to be real, including p53, YY1, c-Ets-1, c-jun, E2F-1, NFκB1, PPAR-alpha, HNF-1A, GR, ER-alpha. Significantly, p53 and Pax-5 binding sites also fell on CpG sites where methylation had a significant impact on HORMAD2 expression. Furthermore, binding sites for RXR-alpha, ER-alpha and TCF-4E were also in close proximity with important CpG sites.

p53 is an important DNA-damage response protein which activates DNA-repair pathways and stalls cell-cycle checkpoints (154), however p53 expression was inversely correlated with HORMAD2 expression **(Appendix C)**. Pax-5 has been identified as a tumour-suppressor in HCC through activation of p53 and p21 signalling (226). Additionally, RXR-alpha forms a heterodimer with PPARα to stimulate genes involved in fatty acid oxidation (227). Importantly, PPARα is a target of adiponectin, and is protective from steatosis (228). RXR-alpha and PPARα expression were indeed positively correlated with HORMAD2, supporting a role for their regulation of HORMAD2 **(Appendix C)**. The oestrogen receptor, ER-alpha (ESR1) also had a predicted binding site close to a significant methylation site, and additionally showed positive correlation with HORMAD2 expression **(Appendix C)**(rho = 0.534, p = $1.80x10^{-32}$).

In all, there is evidence to support that HORMAD2 is regulated by a number of transcription factors, potentially including p53, Pax5, the RXR-alpha/PPAR-alpha complex, and the oestrogen receptor 1. Methylation at the binding sites of these key transcription factors is an important factor in the regulation of HORMAD2 expression, and the high levels of promoter methylation in HCC tumour samples may be responsible for the lack of correlation between some of the transcription factors and HORMAD2 expression. Additionally, microRNAs likely also play an important role in HORMAD2 regulation, particularly via the LOC255167/hsa-mir-182 axis.

Molecular binding partners

The only published binding partner of HORMAD2 is HORMAD1, and it is very likely that HORMAD2 also interacts with itself to form a dimer or potentially an oligomer (36). However, the literature also suggests interactions with DNA repair enzymes such as ATR, BRCA1 and TOPBP1, and TRIP13 (33, 34), however none of these interactions have been identified experimentally. Furthermore, HORMAD2 is known to localise at asynaptic chromosomes in meiosis, but it is not yet known whether this

localisation is mediated by recognition of the DNA itself, of histones, or of other DNA or histone-interacting proteins.

To further elucidate the binding partners of HORMAD2, multiple approaches were taken. First, online databases with results from high-throughput protein-protein interactions studies were searched for HORMAD2 interactions. Hits for HORMAD2 from these databases included NSD2/WHSC1/MMSET (henceforth, NSD2), SYCP2, and ESR2. NSD2 is an oncogenic methyltransferase which promotes proliferation, DNA repair and invasion through histone methylation. NSD2 performs H4K20 methylation of histones at sites of DNA damage, but also directly methylates PTEN, in both cases to allow recognition by 53BP1 via its methylated protein reading activity (217, 229, 230). It is possible that methylation of HORMAD2 by NSD2 could play a role in its recruitment to DNA damage sites, however this conclusion would require far more evidence. SYCP2 functions in meiotic synaptonemal complex formation, and as such its interaction with HORMAD2 is unsurprising, and unlikely to be important in HCC. Finally, the presence of a potential direct interaction with the oestrogen receptor ESR2 is surprising, and particularly interesting given the presence of ESR1 binding sites in the HORMAD2 promoter, and the positive correlation of these genes. Oestrogen and its receptors have been studied extensively in HCC as a potential reason for the sex unbalance in HCC risk. Confusingly however, there appears to be equal evidence to support oestrogen as both a HCC promoter and suppressor, though its role in HCC appears to be related to inflammation and regulation of proliferation (231).

Second, the primary structure (amino acid sequence) was analysed for protein features which suggest molecular interactions. A nuclear localisation signal is present within the C-terminal tail of the protein. This was expected as HORMAD2 is known to be a nuclear protein. Additionally, two S/T-Q motifs are present in the HORMAD2 sequence, which are the sites of phosphorylation by PIKK family members such as ATM, ATR and DNA-PK. This is consistent with previous reports of phosphorylated forms HORMAD2 (36). Novel predicted features of the HORMAD2 amino acid sequence included recognition sites for MAPK and the BRCT domain of BRCA1 and other DNA-repair enzymes. The BRCT interaction peptide is of particular interest, as HORMAD2 has been implicated in the BRCA1-dependant recruitment of ATR to chromosomes (33). However, closer analysis of the BRCT interaction domain shows that while they fit the pSXXF motif, the lack of a proline residue at the +1 position makes this an unlikely BRCT interaction domain (232). Finally, docking motifs were identified for two different common de-ubiquitination enzyme domains, the meprin and TRAF homology (MATH) domain and the ubiquitin-like (UBL) domain.

Thirdly, a tertiary structure of the protein was created using homology-based structural modelling from SWISS-MODEL. This model was then analysed for predicted protein function using the EMBL-EBI tool, “ProFunc”. Unfortunately, this returned no significant results.

Thus, based on the results of previous high-throughput experiments and the use of web-based bioinformatics tools, some more hints towards HORMAD2's function and regulation can be uncovered. For example, HORMAD2 may interact with NSD2, which suggests that it could be directly methylated which would allow recognition by 53BP1. Additionally, the presence of S/T-Q motifs which are likely to be phosphorylated, and potentially a motif for binding with BRCT domains, HORMAD2 is likely to fit within a DNA-damage cascade. Additionally, repeated evidence of oestrogen receptor interaction in the form of both transcription factor activity and direct protein-protein interactions suggests that oestrogen could play some role in the regulation of HORMAD2 activity, though this is confusing given that HORMAD2 clearly plays a more significant role in male biology. Finally, the presence of MAPK-interacting peptides is further evidence to support a role for HORMAD2 in cell-cycle regulation. Of course, each of these points of evidence are weak on their own and could be the result of non-specific hits in high-throughput experiments, however they are nonetheless useful stepping stones towards elucidating the function of a novel gene.

Conclusion

The findings of this investigation suggest that HORMAD2 has previously unreported function in the liver, and its role may be HCC-protective. HORMAD2 expression is frequently lost in HCC due to hypermethylation of its promoter, which blocks bindings sites for P53, Pax-5, RXR-alpha/PPAR-alpha, and the oestrogen receptor 1. Additionally, HORMAD2 activity may be further regulated by miRNAs, particularly miR-182, as well as ubiquitin-mediated degradation. Pathway correlations suggest that HORMAD2 activity is responsive to lipid and cholesterol metabolism-related oxidative stress and subsequent DNA damage, and that its function is in the negative regulation of cell cycle processes. The findings in this report are heavily limited by an inability to distinguish correlation and causation, as well as the directionality of effects. Thus, the proposed functions are only preliminary, and are intended to guide experimental studies which will validate these findings.

Chapter 4: *In vitro* Characterisation of HORMAD2 in HCC

4.1 Introduction

HORMAD2 is a meiotic protein involved in the synapsis checkpoint, however its expression in the liver is associated with HCC and its function here remains to be elucidated. *In silico* data supports a role for HORMAD2 in HCC as a tumour-suppressor, and the data suggests that it may be a negative regulator of the cell-cycle, and that it may function in the DNA-repair pathway or simply in response to oxidative damage. HORMAD2's function has never been investigated *in vitro* in the liver, and thus there exists an important gap in knowledge regarding HORMAD2s influence on hepatocytes.

4.2 Results

4.2.1 HORMAD2 Expression Response to Fructose Treatment

Given the previously observed effect of high fructose diet on HORMAD2 expression *in vivo*, HORMAD2 expression in response to fructose was investigated *in vitro* in Huh7 cells. Huh7 cells were cultured in glucose-free media supplemented with either 5mM Glucose or 5mM Fructose alongside cells in normal high-glucose media as a control. After 24 hours, fructose caused a small decrease in HORMAD2 expression, and this effect was stronger after 48 hours (**Figure 4.1**).

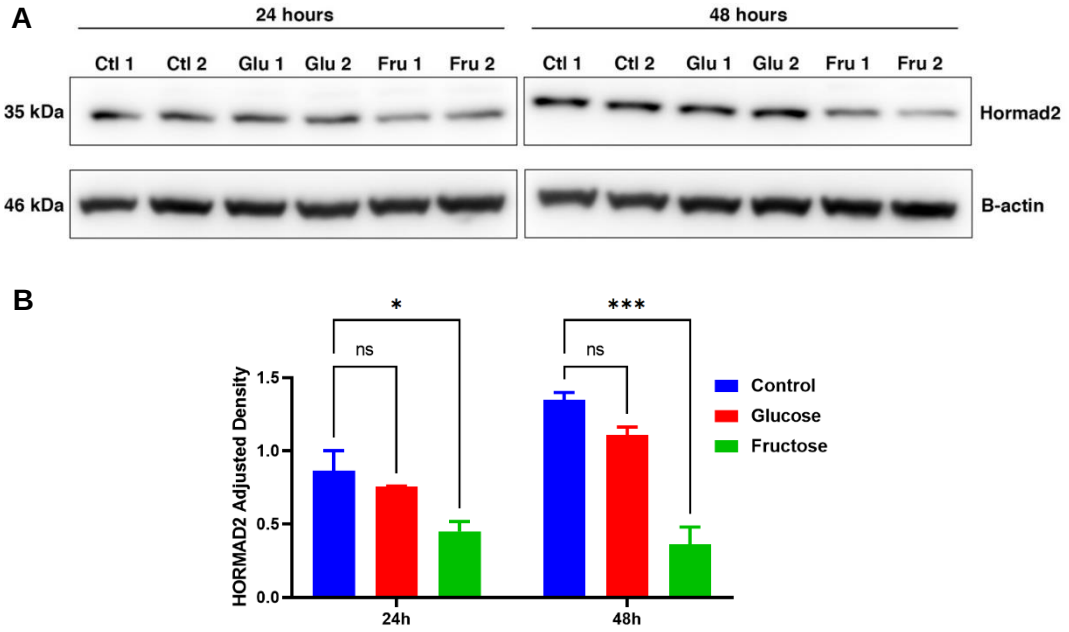

Figure 4.1: Hormad2 expression in glucose- and fructose-treated HCC. **A)** Western blot analysis of HORMAD2 protein expression in Huh7 cells grown in duplicate in control media, media supplemented with glucose, and media supplemented with fructose. **B)** Densitometry of western blot analysis, HORMAD2 density adjusted to beta-actin density. *, $p < 0.05$; ***, $p < 0.001$ (Two-Way ANOVA, Tukey). Ctl, control Huh7; Glu, glucose; Fru, fructose.

4.2.2 Constitutive HORMAD2 Overexpression in HCC Cell Lines

Initially, attempts were made to constitutively overexpress HORMAD2 in HCC cell lines using the pLEGFPN1 plasmid, which is designed for either transient transfection or retrovirus generation in a packaging cell line. This plasmid drives gene expression under the CMV promoter.

The following expression vectors were created using standard cloning techniques:

- **pLN1:** The pLEGFPN1 plasmid with the GFP sequence removed to express no protein but provides G418 resistance. Used as a negative control. Provided by Lionel Hebbard.
- **pL-HORMAD2-FLAG:** The pLEGFPN1 plasmid with the GFP sequence replaced by HORMAD2 with a FLAG tag and G418 resistance.
- **pL-HORMAD2-uvGFP:** The pLEGFPN1 plasmid with the GFP sequence replaced by HORMAD2 fused to the uvGFP sequence from pIM013-uvGFP and G418 resistance.

The plasmid sequences were checked via PCR and sequencing (data not shown).

Initially, retrovirus was created for these constructs in HEK293GP cells using via Lipofectamine 3000 transfection, and virus-containing media was used to transduce the construct into PLC/PRF/5 and Huh7 cell lines. Successfully integrated cells were selected with G418. However, for the Huh7 cell line, only the pLN1 cells continued to divide after selection. In the PLC/PRF/5 cells, protein lysates showed no appreciable change in HORMAD2 expression, and did not present with multiple bands as expected due to the fusion tags (**Figure 4.2A**). Additionally, PLC/PRF/5 pL-HORMAD2-uvGFP cells were not fluorescent under a fluorescence microscope (data not shown).

As an alternative, transient HORMAD2 overexpression was attempted using pL-HORMAD2-FLAG and Lipofectamine 3000. Protein lysates prepared 48 hours after transfection showed no evidence of HORMAD2 over-expression (**Figure 4.2B**)

Finally, stable transfection was attempted via transfection with Lipofectamine 3000 and selection with G418 in Huh7, PLC/PRF/5 and Hep3B. In all cases, after several months of culture, HORMAD2-transfected cell lines failed to grow.

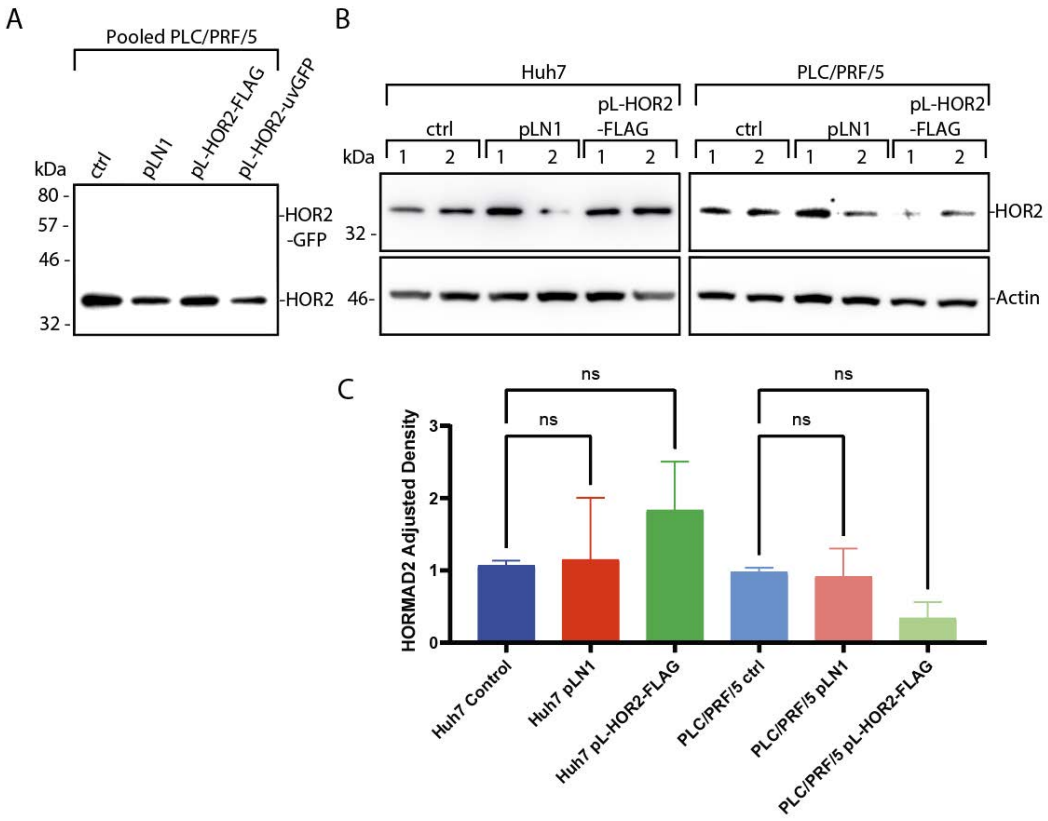

Figure 4.2: Unsuccessful constitutive overexpression of HORMAD2. **A)** Retrovirus was used to transduce PLC/PRF/5 cells with pL-HORMAD2-FLAG and pL-HORMAD2-uvGFP, as well as pLN1 as a control. There was no evidence of successful overexpression. **B)** The same constructs were then used to transiently overexpress HORMAD2 in duplicate for 48 hours. This again caused no evidence of successful overexpression. **C)** Densitometry of western blot analysis, HORMAD2 density adjusted to beta-actin density (One-Way ANOVA, Sidak). HOR2, HORMAD2; GFP, green-fluorescent protein.

4.2.3 Inducible HORMAD2 Overexpression in HCC Cell Lines

Given the previous failure of constitutively overexpressing HORMAD2 in HCC cell lines, a new construct was created for doxycycline-inducible HORMAD2 overexpression. Additionally, supply of the HORMAD2 antibody had diminished, and resupply was inaccessible due to short supply from the manufacturer and shipment delays in the COVID-19 pandemic. Therefore, the new construct was created with a 6HIS tag for western blot purposes.

Using standard cloning procedures, the following plasmid vectors were produced.

- **PCW(Empty):** The PCW57.1 plasmid that does not produce any protein, but provides puromycin resistance. Negative control.
- **PCW-HORMAD2:** The PCW57.1 plasmid with a HORMAD2 sequence amplified from pL-HORMAD2-FLAG and an added stop codon before the 6-HIS tag. Also provides puromycin resistance.

- **PCW-HORMAD2-6HIS:** The PCW57.1 plasmid with a HORMAD2 sequence amplified from pL-HORMAD2-FLAG with produces transgene with a 6HIS fusion tag. Also provides puromycin resistance.

The plasmid sequences were checked via PCR and sequencing (data not shown).

For each of the PCW57.1 constructs, lentivirus was produced in HEK293T cells via $CaCl_2$ transfection. Virus-containing media was then collected and concentrated via ultracentrifugation. Hep3B and Huh7 cell lines were transduced with these viral constructs and selected using puromycin, based on puromycin minimum toxic concentrations identified previously (data not shown). PCR amplification of genomic DNA from infected Hep3B cells indicated successful viral integration (**Figure 4.3A**). Doxycycline induction of the HORMAD2-6HIS protein was successful as indicated by western blot against the 6HIS tag (**Figure 4.3B, G**), however blotting against HORMAD2 demonstrated that the resulting increase in overall HORMAD2 expression was only small (**Figure 4.3E and F**).

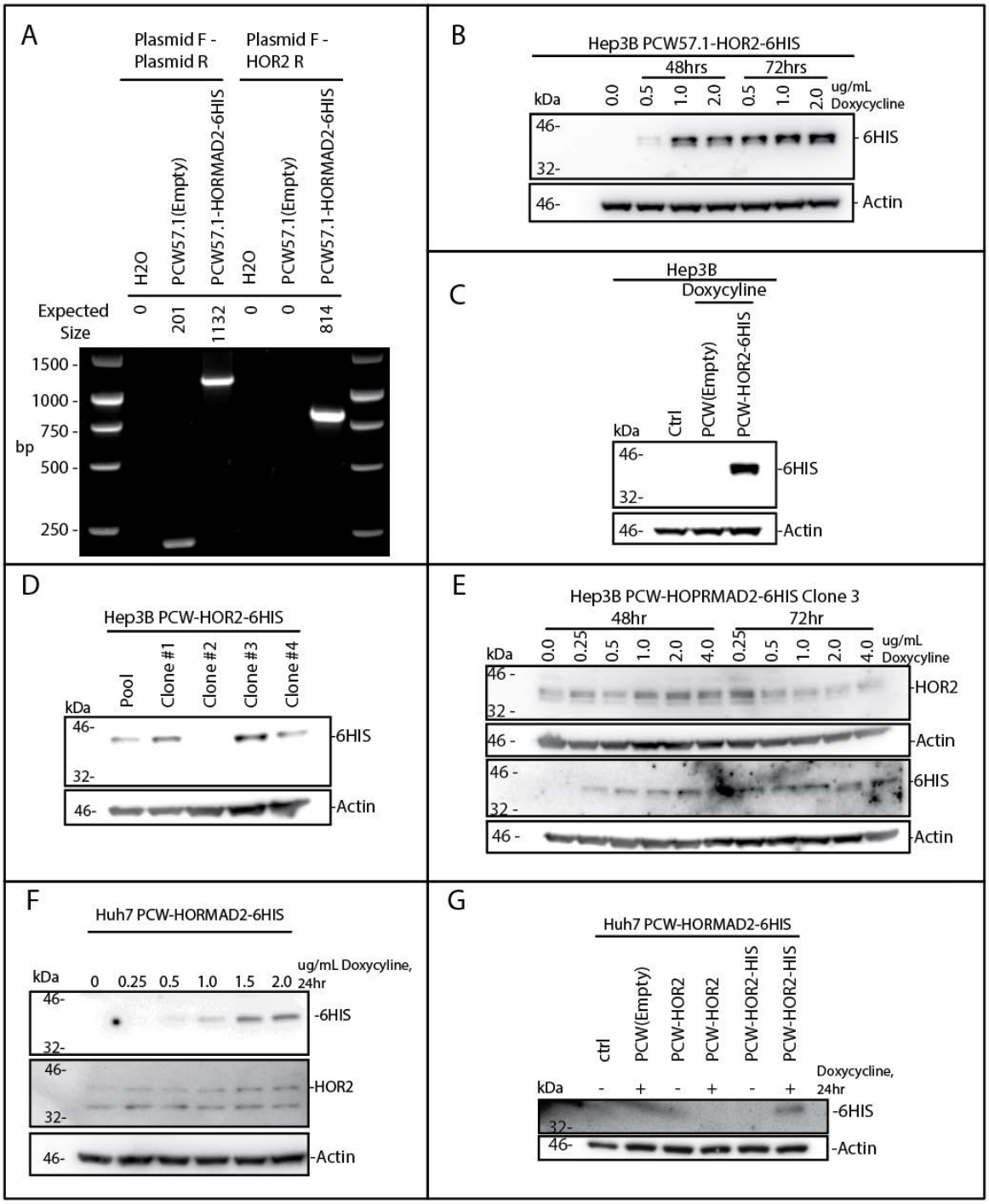

Figure 4.3: Successful inducible expression of HORMAD2-6HIS. Lentivirus was produced for the PCW-HORMAD2-6HIS, PCW-HORMAD2, and PCW(Empty) constructs in HEK293T cells. Hep3B and Huh7 cells were transduced with virus and selected with puromycin. **A)** Genomic DNA was purified from transduced Hep3B cells and primers targeting the PCW57.1 plasmid and the HORMAD2 sequence were used with PCR to confirm viral integration (**Table 2.5**). **B and F)** Modified cells were cultured with increasing concentrations of doxycycline to optimise expression induction. **C and G)** Expression of HORMAD2-6HIS was evaluated after doxycycline-induction in comparison with doxycycline-treated PCW(Empty) cells and untreated control cells. **D)** Individual clones of Hep3B PCW-HORMAD2-6HIS cells were generated by dilution-selection of individual cells. HOR2, HORMAD2;

4.2.4 HORMAD2 shRNA Knockdown Construct

Knockdown of HORMAD2 expression was performed using short-hairpin RNA (shRNA) constructs. In total, constructs with five different HORMAD2-targeting sequences were purchased.

- **sh Scrm:** Produces a non-targeting "Scram" shRNA as a negative control and provides puromycin resistance.
- **sh #1-5:** MISSION shRNA plasmids with five different HORMAD2-targeting sequences. Also provides puromycin resistance.

For each of the shRNA constructs, lentivirus was produced in HEK293T cells via $CaCl_2$ transfection. Virus-containing media was then collected and concentrated via ultracentrifugation. Huh7 cell lines were transduced with these viral constructs and selected using puromycin, based on puromycin minimum toxic concentrations identified previously (data not shown). Western blot showed that sh-HORMAD2 #2 #3 had the strongest effect on HORMAD2 expression (**Figure 4.4**).

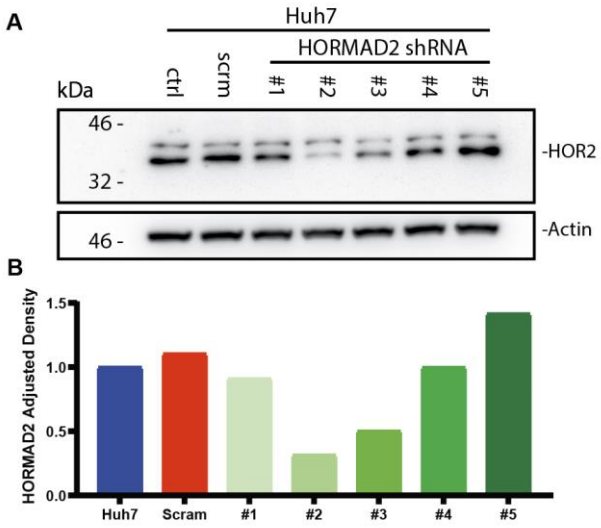

Figure 4.4: Successful knockdown of HORMAD2 expression in Huh7 cells. Five lentiviral shRNA constructs targeting HORMAD2 were created in HEK293 cells and evaluated in comparison to a non-targeting 'Scram' shRNA virus. Transduced Huh7 cells were selected with puromycin then cultured without puromycin for several days before **A)** HORMAD2 concentration was evaluated via western blot of protein lysates. **B)** Densitometry of western blot analysis, HORMAD2 density adjusted to beta-actin density Ctrl, control Huh7; scrm, non-targeting Scram virus; shRNA, short-hairpin RNA; HOR2, HORMAD2

4.2.5 Nuclear Localisation of HORMAD2

Fractionation protein lysates were performed to independently extract nuclear, cytoplasmic and membrane lysate fractions in order to identify localisation of HORMAD2-6HIS protein within the Hep3B HCC cell line (**Figure 4.5**). HORMAD2-6HIS was found to be expressed most strongly in the nucleus of these cells, however cytoplasmic expression was also detected. HORMAD2-6HIS was not obviously associated with insoluble membranous fraction of the lysate.

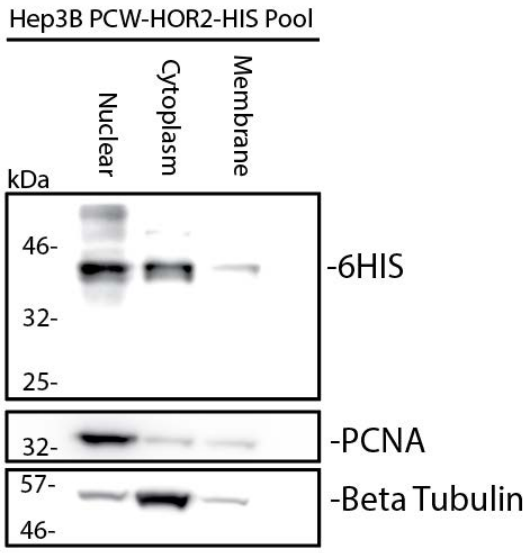

Figure 4.5: Nuclear localisation of HORMAD2. Hypotonic nuclear fractionation was performed on Hep3B PCW-HORMAD2-6HIS cells after induction with doxycycline. Nuclear, cytoplasmic and membrane (insoluble) fractions were evaluated for HORMAD2 expression by blotting for the 6HIS tag. PCNA was used a nuclear marker, beta tubulin was used as a cytoplasmic marker.

4.2.6 HCC Cell Proliferation

Previous *in silico* data suggested that HORMAD2 may affect the cell cycle. Additionally, unpublished observations indicated that HORMAD2-overexpressing cells grew slower, and HORMAD2-knockdown cells appeared to proliferate faster. Thus, proliferation assays were used to evaluate the effect of HORMAD2 expression on the modified Huh7 cells **(Figure 4.6)**. Neither overexpression nor knockdown of HORMAD2 produced a consistent and significant effect on cell growth greater than the variation between control cells and those treated with non-targeting vectors.

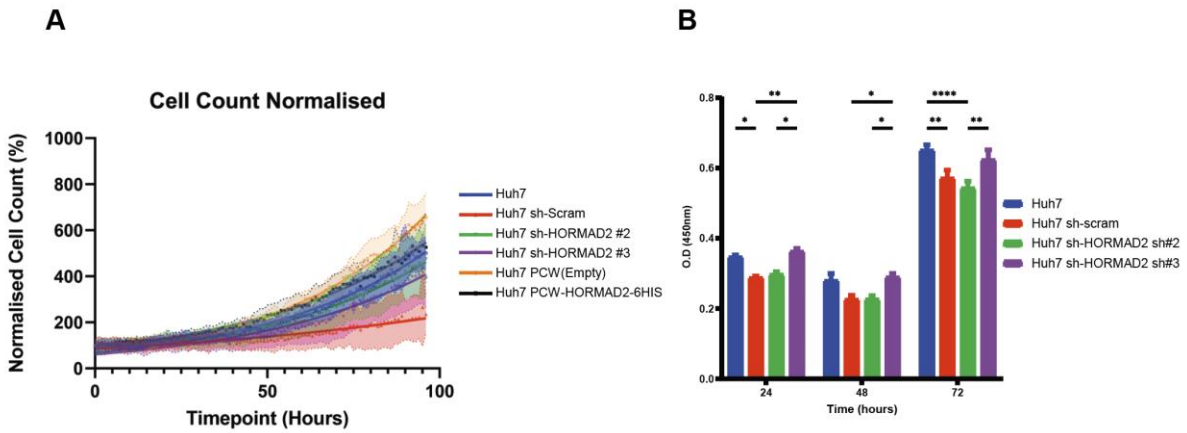

Figure 4.6: HCC cell proliferation after altered HORMAD2 expression. A) Huh7, Huh7 sh-Scram, Huh7 sh-HORMAD2 #2, Huh7 sh-HORMAD2 #3, Huh7 PCW(Empty), Huh7 PCW-HORMAD2-6HIS cells were cultured for 96 hours and cell count was analysed once every hour at 10 positions (n=10) using a Holomonitor (Phi). **B)** Huh7, Huh7 sh-Scram, Huh7 sh-HORMAD2 #2, Huh7 sh-HORMAD2 #3 (n=8 per cell line per timepoint) was evaluated after 24, 48 and 72 hours using a BrdU assay (Roche). Data is shown as mean and SEM. Statistical significance was determined with two-way ANOVA and Tukey's multiple comparisons test. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

4.2.7 HCC Cell Survival Survival Using MTT Assays

4.2.7.1 Chemotherapeutic Toxicity

MTT assays were used to evaluate whether changes in HORMAD2 expression affected resistance to two chemotherapeutic agents. The modified Hep3B and Huh7 cell lines were cultured with increasing concentrations of Sorafenib and Cisplatin. Overall, the effects of these agents on the cells were variable between the cells, and changes in HORMAD2 expression did not cause a change in survival greater than the variation observed between the untreated cells and those treated with empty or non-targeting vectors **(Figure 4.7)**. At low concentrations of Sorafenib (1-4μM), overexpression of HORMAD2 caused a small but significant decrease in cell survival ($p < 0.05$), however on repeat experiments this effect was not reproduced (data not shown).

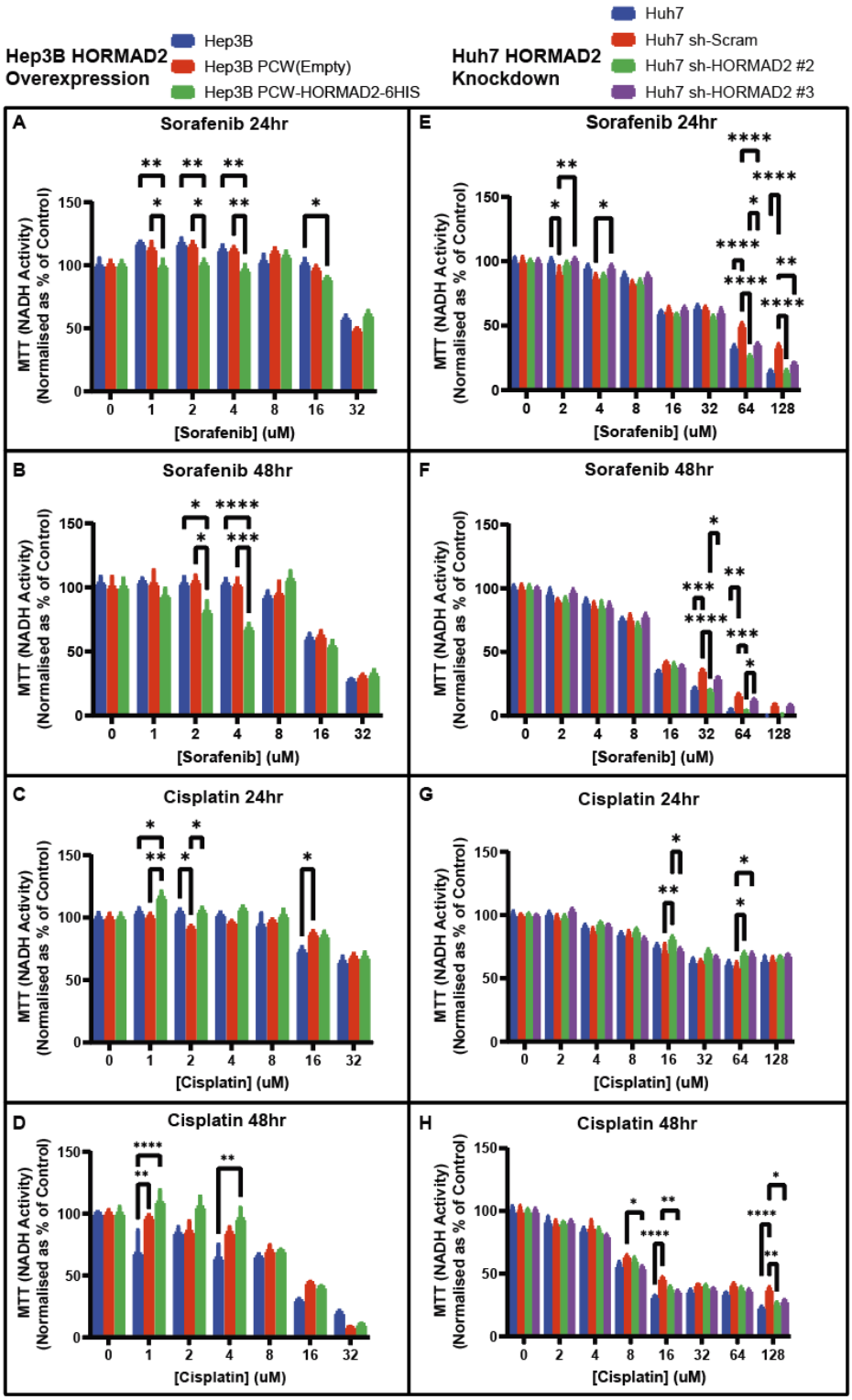

Figure 4.7: Hep3B HORMAD2-ovexpression and Huh7 HORMAD2-knockdown response to chemotherapy. Hep3B, Hep3B PCW(Empty), or Hep3B-PCW-HORMAD2-6HIS (**A-D**) (n = 4 per cell line per treatment) and Huh7, Huh7 scram, or Huh7 Hormad2 knockdown cell lines (sh-HORMAD2 #2 and sh-HORMAD2 #3) (**E-F**) (n = 5 per cell line per treatment) were treated with Sorafenib (**A,B,E,F**) and Cisplatin (**C,D,G,H**) for 24 (**A,E,C,G**) and 48 **(B,F,D,H)** hours, respectively. Cell proliferation was assessed using MTT and absorbance at 540 nm was calculated as a percentage of the average absorbance of non-treated controls for each cell line. Data is shown as mean and SEM. Statistical significance was determined with two-way ANOVA and Tukey's multiple comparisons test. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

4.2.7.2 Oxidative Stress Toxicity

Preliminary data indicated that high fructose diets and fatty liver disease contribute to changes in HORMAD2 expression. Additionally, fructose is toxic to HCC cell lines in culture (**Figure 4.8B,D**). Thus, MTT assays were utilised to investigate the resistance of HORMAD2-manipulated HCC cell lines to H_2O_2 as a model of oxidative stress as well as fructose **(Figure 4.8)**. HORMAD2 overexpression in the Hep3B cell line slightly improved the survival of cells after treatment with H_2O_2 and fructose relative to cells transfected with an empty vector, but the difference was not significant when compared to non-transfected cells. Thus these changes were within the overall variability seen in the control cell lines. The same was true for the knockdown of HORMAD2 in Huh7 cells using sh-HORMAD2 #3 in response to fructose treatment, and again no significant difference was seen in HORMAD2-knockdown cells relative to the untreated control after H_2O_2 treatment. Therefore, in all cases, changes in HORMAD2 expression did not cause a change in survival greater than the variation observed between the untreated cells and those treated with empty or non-targeting vectors.

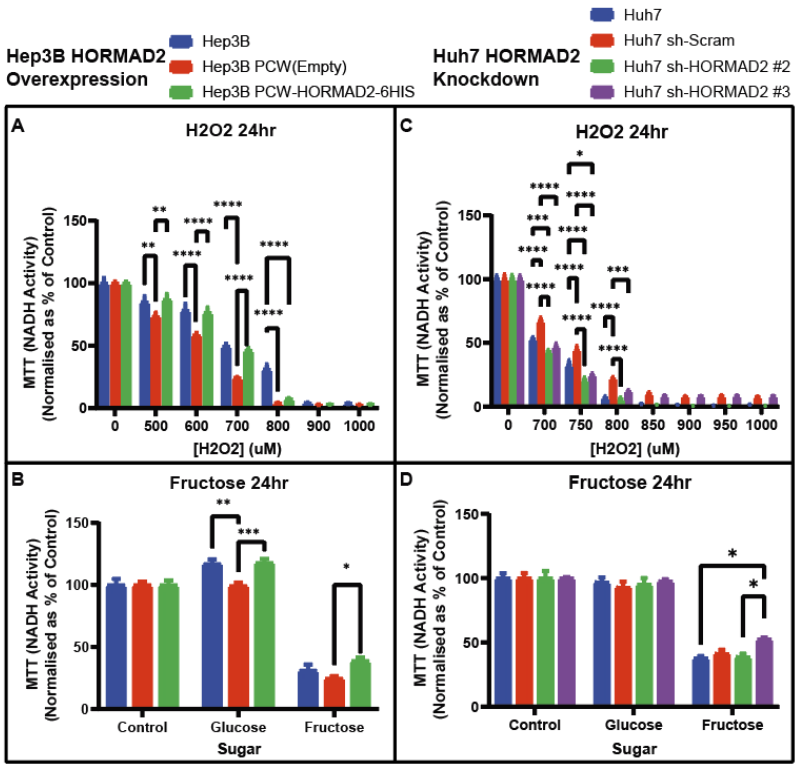

Figure 4.8: Hep3B HORMAD2-ovexpression and Huh7 HORMAD2-knockdown response to oxidative stress. Hep3B, Hep3B PCW(Empty), or Hep3B-PCW-HORMAD2-6HIS (**A,B**) and Huh7, Huh7 scram, or Huh7 Hormad2 knockdown cell lines (sh-HORMAD2 #2 and sh-HORMAD2 #3) (**C,D**) were cultured in normal media, glucose or fructose-supplemented media (**B,D**) or increasing concentrations of H_2O_2 (**A,C**) for 24 hours. Sample sizes per cell line and per treatment were as follows: Hep3B H_2O_2, n = 4; Hep3B fructose, n = 5, Huh7 H_2O_2, n = 5; Huh7 fructose, n = 8. Cell proliferation was assessed using MTT and absorbance at 540 nm was calculated as a percentage of the average absorbance of non-treated controls for each cell line. Data is shown as mean and SEM. Statistical significance was determined with two-way ANOVA and Tukey's multiple comparisons test. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

4.2.8 Radiation Sensitivity

To determine the sensitivity of HORMAD2-knockdown HCC cells to radiation-induced DNA damage, Huh7, Huh7 scram and Huh7 HORMAD2-knockdown (sh-HORMAD2 #2 and sh-HORMAD2 #3) were exposed to a low, medium, and high dose of gamma irradiation (1.8 Gy, 5.4 Gy, and 9 Gy, respectively). Tumoursphere assay was used to assess cell stemness in non-irradiated cells and irradiated cells **(Figure 4.9).** Without radiation, Huh7 sh-HORMAD2 #2 cells produced significantly less tumourspheres than unmodified cells, however there was no significant difference with the sh-Scram-treated cells or the other knockdown construct. Overall, HORMAD2 expression did not noticeably affect tumoursphere number **(Figure 4.9A)**. However, it was noted that, consistently, HORMAD2-knockdown led to the formation of considerably larger and aggregative tumourspheres. Indeed, even after high dose radiation (9 Gy) the ability to form large and aggregated tumourspheres persisted in the HORMAD2-knockdown cells **(Figure 4.9B)**.

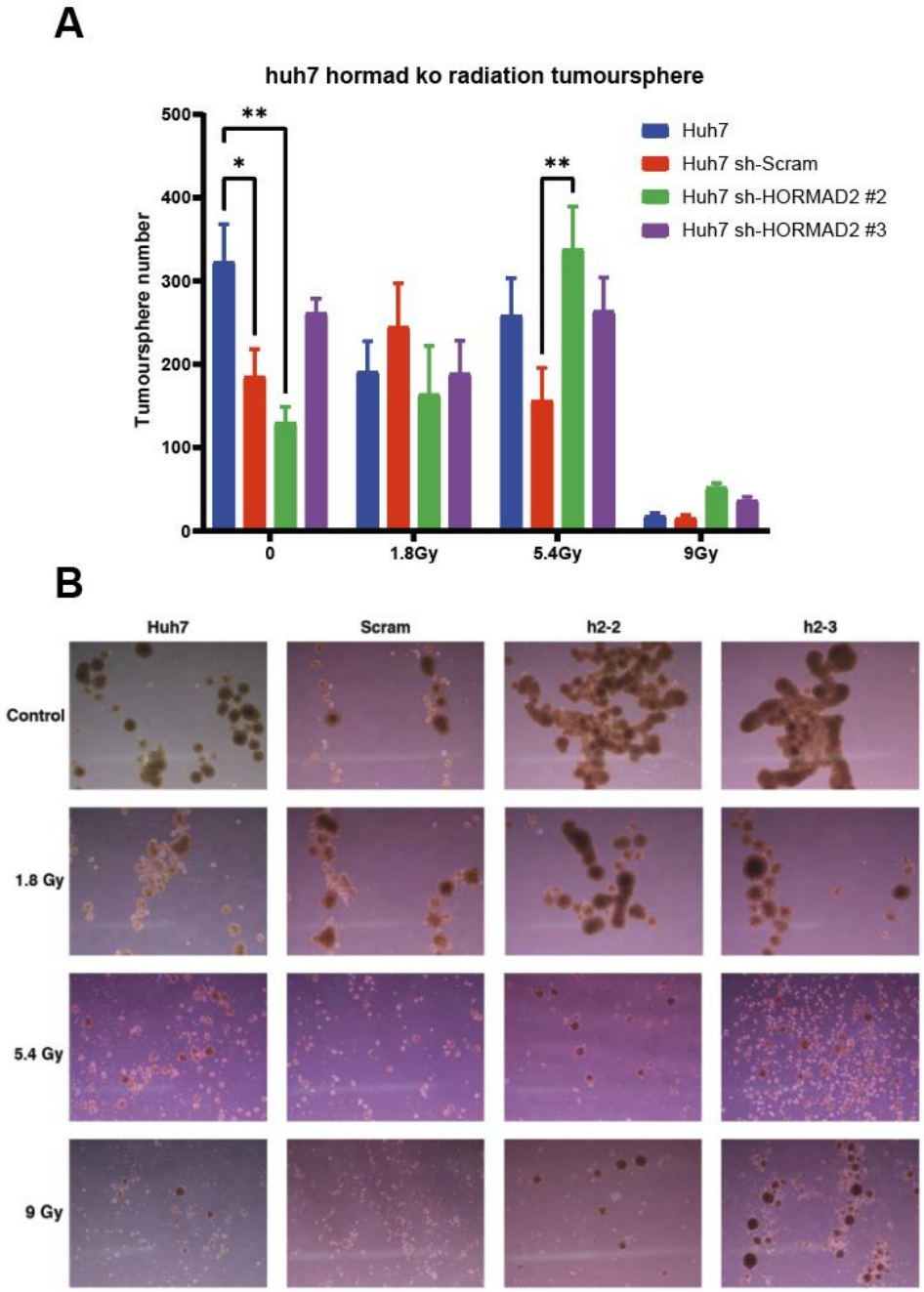

Figure 4.9: Huh7 HORMAD2 knockdown tumoursphere formation. Huh7, Huh7 sh-Scram, or Huh7 HORMAD2 knockdown cell lines (sh-HORMAD2 #2 and sh-HORMAD2 #3) were treated with a low (1.8 Gy), medium (5.4 Gy), or high dose (9 Gy) of radiation using an Elekta linear accelerator and tumoursphere formation was compared to non-irradiated cells (n = 3 per cell line per treatment). **A)** Tumourspheres were quantified after 7 days and, **B)** representative sphere photos are shown. Data is shown as mean and SEM. Statistical significance was determined with two-way ANOVA and Tukey's multiple comparisons test. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

To examine the ability of HORMAD2-knockdown HCC cells to self-proliferate after radiation-induced DNA damage, a colony formation assay was performed using cells treated with various doses of radiation as above **(Figure 4.10)**. The plating efficiency (PE) was determined as the number of colonies formed per number of cells seeded. The surviving fraction (SF) was determined as the PE of each cell line after radiation multiplied by the PE of their non-irradiated counterparts. Before and after radiation,

Huh7 sh-HORMAD2 #3 had the highest PE and it was significantly higher than sh-Scram treated cells before radiation, however it was not significantly higher than unmodified Huh7 cells at any radiation dosage. Both the sh-Scram-treated cells and sh-HORMAD2 #2 had consistently lower PE than unmodified Huh7s or those treated with sh-HORMAD2 #3. The two HORMAD2 knockdown cell lines showed varying abilities to form colonies after DNA damage, and Huh7 scram were not comparable to normal Huh7. No colonies were formed for any cell line after 9 Gy irradiation.

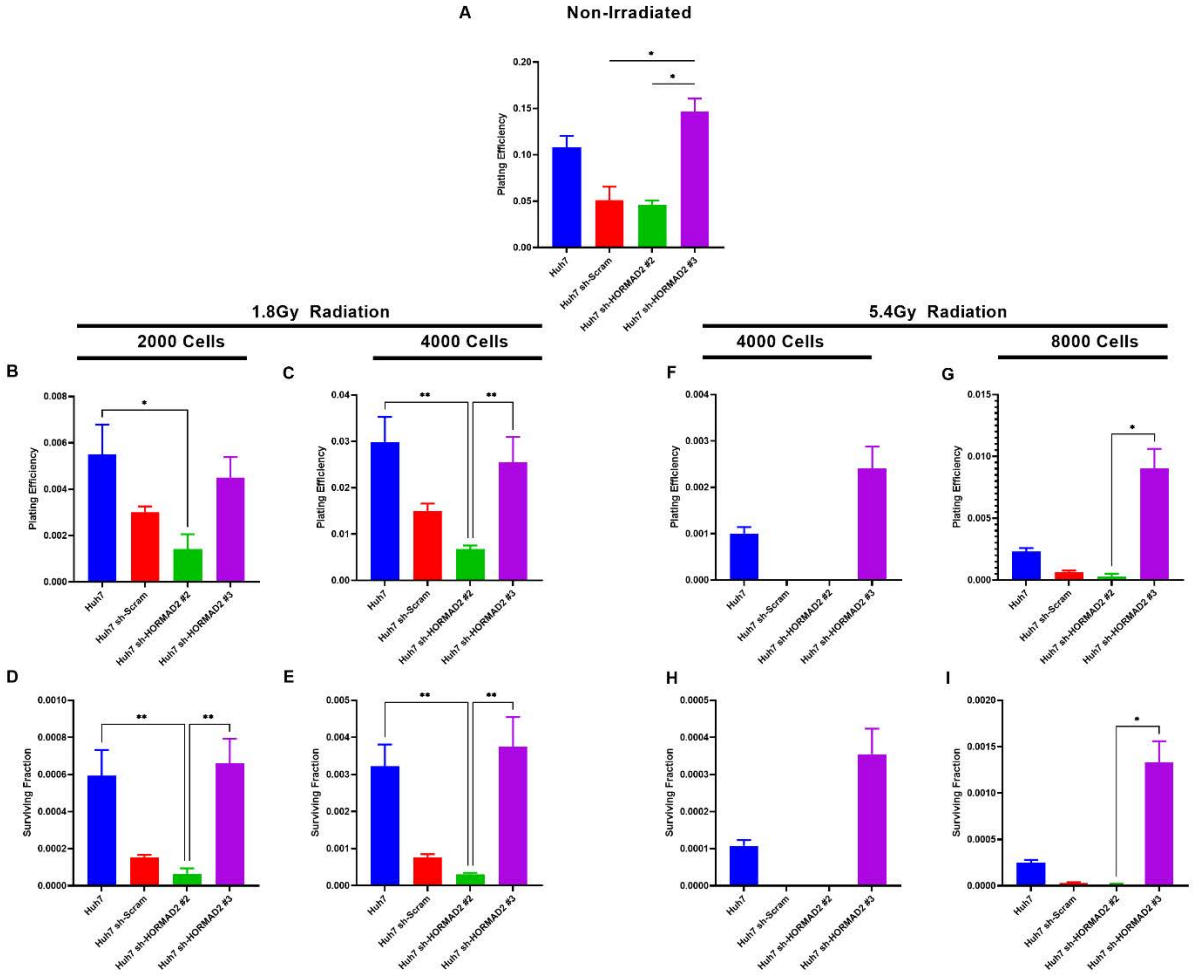

Figure 4.10: HCC cell colony formation after HORMAD2 knockdown and radiation. A) The plating efficiency was determined for Huh7, Huh7 sh-Scram, or Huh7 HORMAD2 knockdown cell lines (sh-HORMAD2 #2 and sh-HORMAD2 #3) by counting the number of colonies formed after 14 days from 1000 seeded cells (n = 5-8 per cell line per treatment). **B,C,F,G)** Plating efficiency was calculated, as well as **D,E,H,I)** surviving fraction as a product of plating efficiency before and after radiation at **B-E)** 1.8 Gy and **F-I)** 5.4 Gy. Sample sizes vary due to extensive cell death after radiation, 1.8Gy n = 4-6 per cell line per plating density, 5.4Gy n = 1-3 per cell line per plating density. Data is shown as mean and SEM. Statistical significance was determined with one-way ANOVA/Kruskal-Wallis test and Dunn's multiple comparisons test. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

4.2.9 HORMAD2 Effect on DDR Pathways

Huh7 control and modified cells were irradiated with various doses of radiation as above and the phospho-activation of ATM, ATR and p53 was evaluated via western blot of protein lysates prepared within 50±5 minutes (Average ± SD) of radiation treatment **(Figure 4.11)**. As expected, increasing radiation dosage causes increased phosphorylation of ATM, ATR, and P53. For each of these proteins, staining was evident at both the expected size of each protein, as well as at a number of smaller sizes indicating degradation of the proteins. This was likely also a result of extensive non-specific staining. There was no apparent difference in activation of these pathways between the HORMAD2-knockdown cell lines compared with controls.

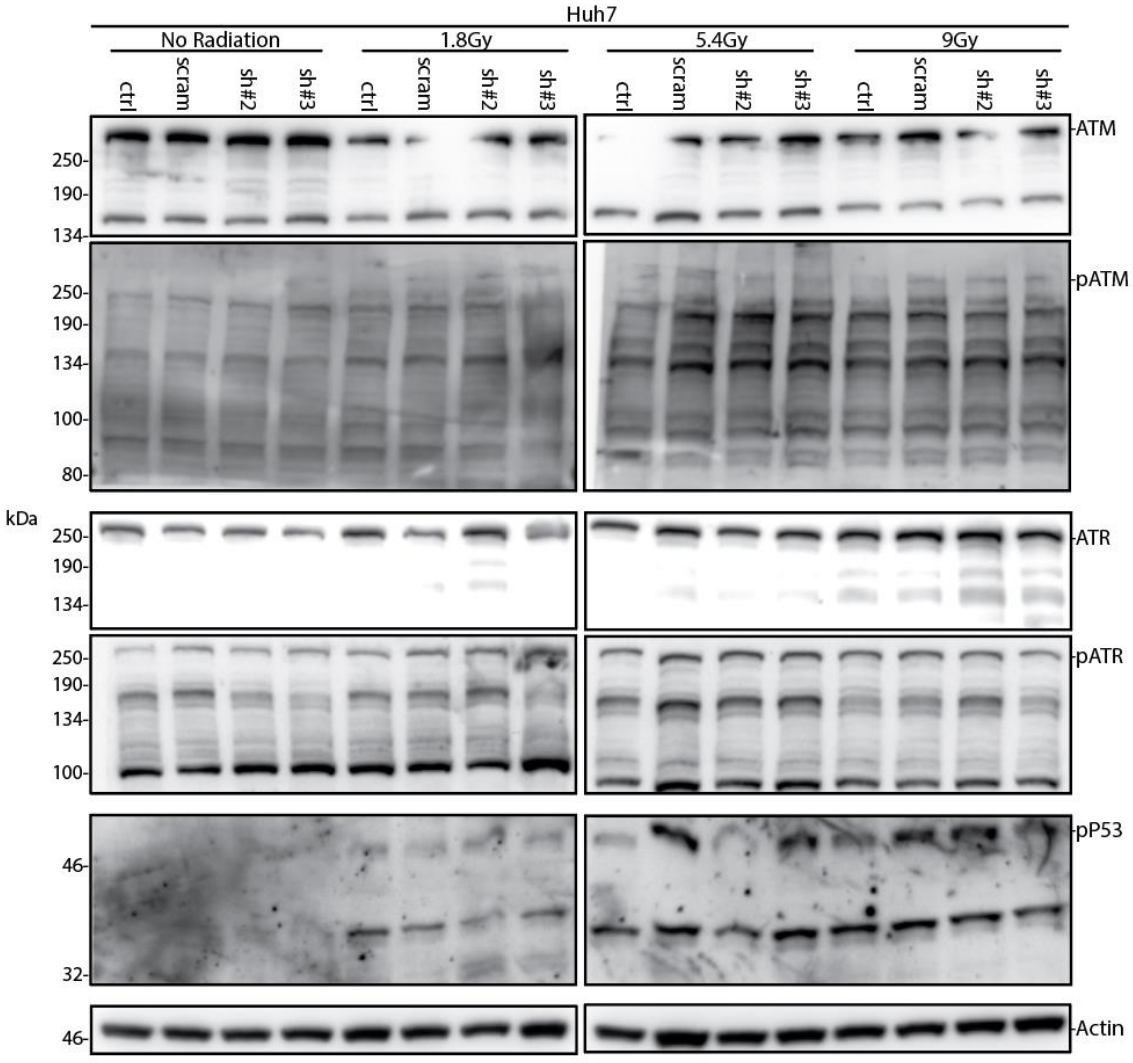

Figure 4.11: Expression of DDR proteins in HORMAD2 knockdown HCC after radiation. Protein expression of ATM, phopho-ATM, ATR, phospho-ATR and phosphor-p53 were determined in Huh7, Huh7 sh-Scram, or Huh7 HORMAD2 knockdown cell lines (sh-HORMAD2 #2 and sh-HORMAD2 #3) exposed to 1.8 Gy, 5.4 Gy and 9 Gy irradiation via western blot and were compared to non-irradiated controls. Ctrl, Control; ATM, Ataxia-Telangiectasia Mutated; ATR, Ataxia telangiectasia and Rad3 related.

4.2.10 HORMAD2 and Overall DNA Damage via Comet Assay

Comet assays were performed on control and modified Huh7 cells to evaluate overall levels of DNA damage 1 hour after 500μM H_2O_2 treatment and 1 hour after 5.4 Gy radiation. After treatment, cells were lysed and imbedded in an agarose gel before being subjected to electrophoresis. The length of the resulting "comets" **(Figure 4.12C)** were used as a measure of overall DNA damage. For the radiation-induced damage, cells (including the undamaged control) were frozen at -80°C for several weeks before the assay was performed, hence the general increase in DNA damage. After 1 hour treatment with H_2O_2 HORMAD2-knockdown cells had significantly shorter comet length ($p < 0.05$, **Figure 4.12A**). Additionally, HORMAD2-6HIS overexpressing cells had shorter comets than normal Huh7s after H_2O_2-treatment, however the PCW(Empty) vector also caused a significant decrease in comet length, suggesting that other factors such as doxycycline-induction may have confounded this finding. 1hr after 5.4 Gy radiation, HORMAD2-knockdown with sh-HORMAD2 #2 also caused a significant decrease in comet length ($p < 0.05$, **Figure 4.12B**), however the decrease caused by sh-HORMAD2 #3 was non-significant ($p > 0.05$, **Figure 4.12B**).

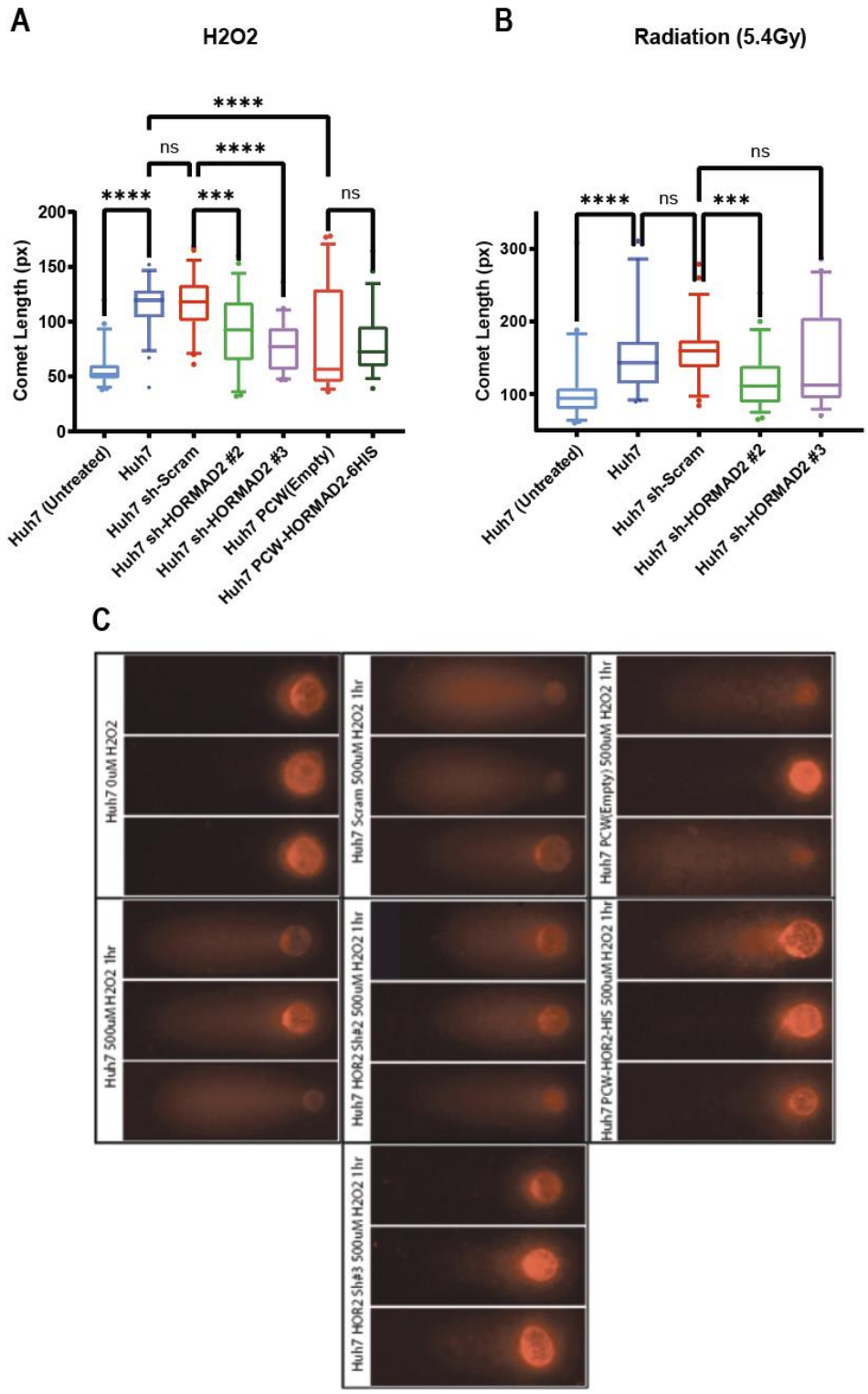

Figure 4.12: Huh7 HORMAD2 knockdown and overexpression comet assays. Huh7, Huh7 scram, Huh7 HORMAD2 knockdown cell lines (sh-HORMAD2 #2 and sh-HORMAD2 #3), Huh7 PCW(Empty) or Huh7 PCW-HORMAD2-6HIS were treated with **A)** 500 μM H_2O_2 in serum-free media or, **B)** 5.4 Gy, radiation using an Elekta linear accelerator. Comet assay was performed using standard techniques on cells **A)** 1hr after treatment or **B)** on cells frozen at -80°C 1hr after treatment. Comet length of 45-50 comets per sample was calculated manually using CometScore 2.0.0.38. **C)** Representative 'comets' are shown. Data is shown as mean and 5-95% percentile. Statistical significance was determined with one-way ANOVA and Tukey's multiple comparisons test. * $p<0.05$; ** $p<0.01$; *** $p<0.001$; **** $p<0.0001$.

4.2.11 Recombinant HORMAD2-GFP

A plasmid was designed and purchased for the production of a recombinant 6HIS-HORMAD2-GFP fusion protein in T7-polymerase-expressing *E. coli* hosts. This plasmid was transformed into KRX, BL21(DE3)RIPL, and Origami B(DE3) strains using standard techniques, and protein expression was induced using rhamnose or 4mM galactose for approximately 72 hours. Recombinant protein was recovered and purified via French-press lysis followed by Ni-IMAC column purification.

The expression and purification of HORMAD2 was compared with GFP protein alone. Firstly, the level of fluorescence in *E. coli* hosts normalised by cell density shows that protein expression was much weaker for HORMAD2 **(Figure 4.13A)**. Secondly, fluorescence microscopy demonstrated that, in comparison with GFP protein alone, HORMAD2 was prone to formation of inclusion bodies, indicative of insoluble protein **(Figure 4.13B)**. Thirdly, SDS-PAGE showed that yields were low due to protein insolubility and proteolysis **(Figure 4.13C)**. Nonetheless, trials of various buffers and *E. coli* strains led to an optimised protocol for HORMAD2 production in KRX cells with phosphate buffers and beta-mercaptoethanol reducing agent **(Figure 4.13D)**. For experimental analysis, protein was concentrated with ammonium-sulphate precipitation and resuspension in a simple 50mM phosphate buffer.

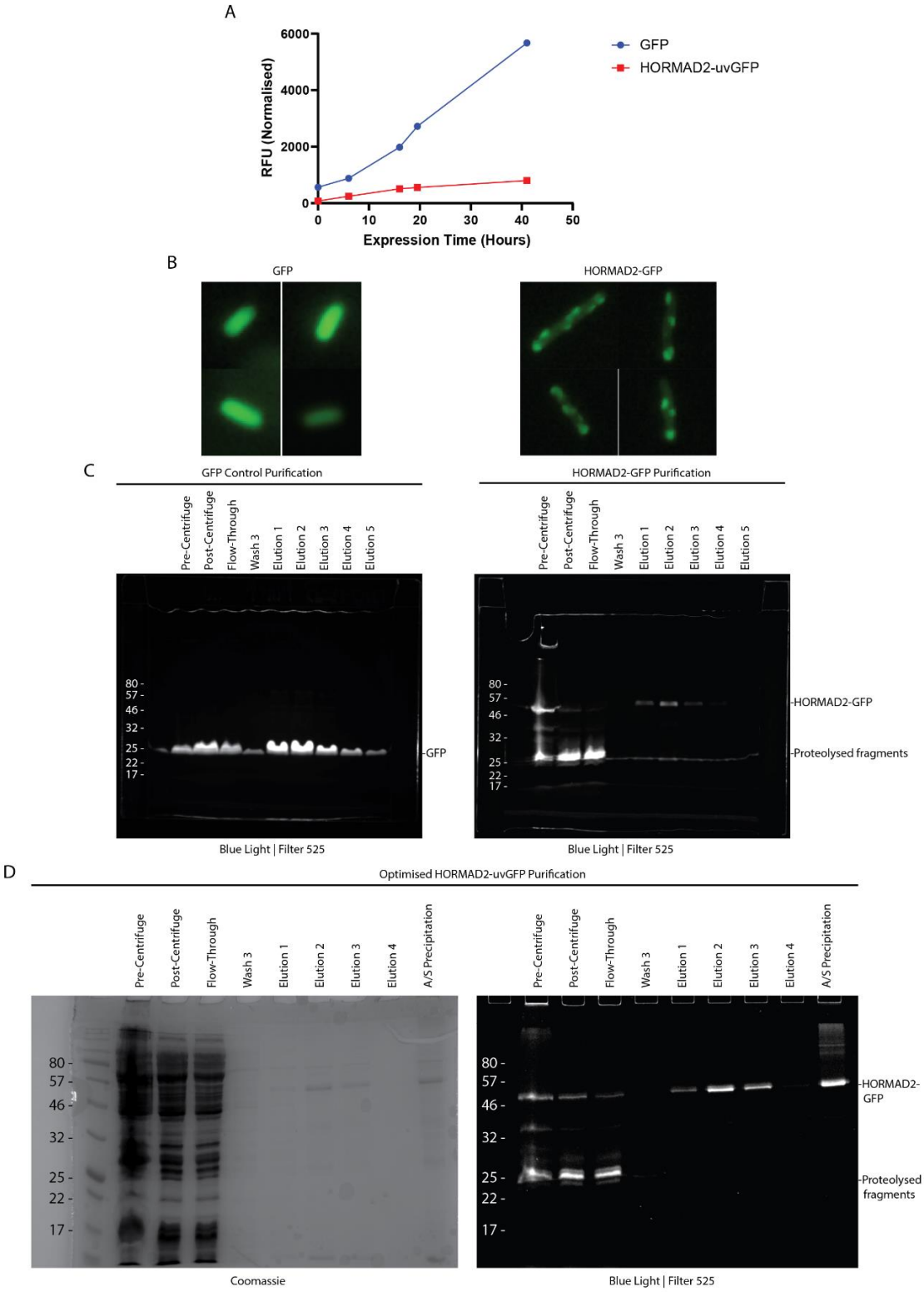

Figure 4.13: Expression of recombinant HORMAD2-uvGFP. A-C) In BL21(DE3)-RIPL *E. coli* hosts, expression was compared between HORMAD2-uvGFP and uvGFP alone. A) Fluorescence of bacterial culture was investigated over time in aliquots of normalised OD_{600} of bacterial culture. B) Fluorescence microscopy was used to investigate formation of inclusion bodies within *E. coli*. C) SDS-PAGE was performed on aliquots throughout purification process and imaged for GFP-fluorescence. D) SDS-PAGE of optimised HORMAD2-uvGFP purification in KRX. RFU, Relative fluorescence units; GFP, Green fluorescent protein; A/S, ammonium sulphate

4.2.12 GFP-EMSA

GFP electrophoretic mobility shift assays (GFP-EMSA) were performed by running various combinations of purified HORMAD2-GFP and DNA samples on a 1% agarose gel after 10-minute incubations at room temperature. Recombinant HORMAD2 protein showed evidence of RNA contamination from the purification process, so HORMAD2-GFP was also treated with RNase prior to a subset of the experiments as indicated. Migration of the human genomic DNA (sheared via 29G syringing) and bovine genomic DNA (restriction enzyme digested) was different due to different methods of DNA shearing. HORMAD2 caused a DNA band shift for both human and bovine genomic DNA, and caused smearing of the short fragment BioO DNA, all of which indicate binding. Little effect was observed on the protein migration under GFP fluorescence (right), however in lanes combined with DNA, GFP fluorescence was evident aggregating within the well of the agarose gel **(Figure 4.14)**.

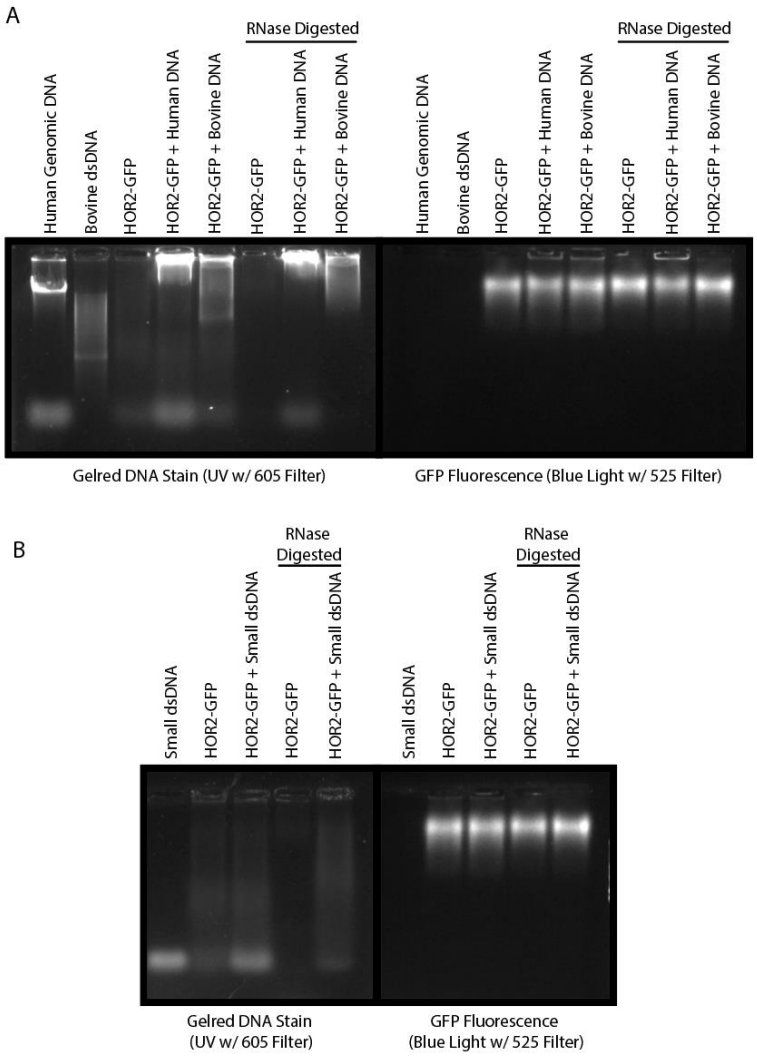

Figure 4.14 GFP EMSA indicating DNA-binding of HORMAD2. Various combinations of HORMAD2-uvGFP recombinant protein and DNA samples (A) human and bovine genomic DNA, B) BioO short dsDNA) were run together or independently on a 1% agarose gel, imaged for GFP-fluorescence and then post-stained with GelRed and imaged again for DNA migration. Where indicated, some samples of HORMAD2-uvGFP were treated with RNaseA for 1 hour prior to combination with DNA samples.

4.2.13 GFP-BASTA

Using a GFP-basted stability assay (GFP-BASTA), the aggregation temperature of HORMAD2-uvGFP was determined to be approximately 69°C (SDS-PAGE: 68.89, Plate Reader: 70.03), while the aggregation of GFP alone was approximately 78°C (SDS-PAGE: 78.11, Plate Reader: 78.16) **(Figure 4.15A-D)**. Recombinant HORMAD2-uvGFP was then combined with $MgCl_2$, $MnCl_2$, ZnAc, Bovine genomic DNA, a combination of those metal compounds with DNA, or ATP and heated to 69°C, followed by centrifugation to observe any changes in fluorescence indicative of a stabilising or destabilising interactions **(Figure 4.15E)**. None of these combinations caused a strong enough shift from 50% remaining fluorescence to indicate a stabilising or destabilising interaction.

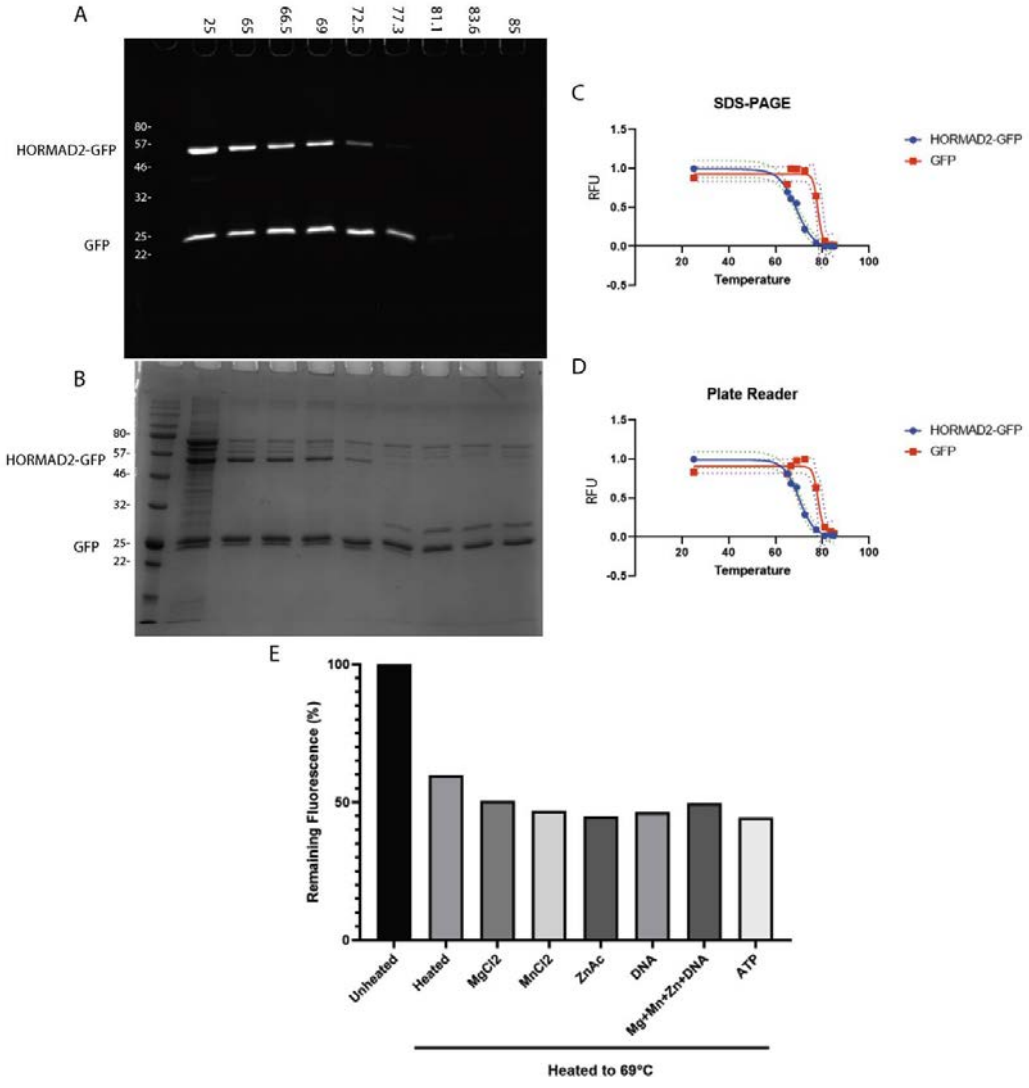

Figure 4.15: GFP-BASTA of HORMAD2 to identify molecular interactions. HORMAD2-uvGFP and GFP alone were heated over 65-85°C and, centrifuged, and run on an SDS-PAGE (**A)** GFP-fluorescence, **B)** Coomassie blue post-staining), and GFP fluorescence was recorded in a plate-reader. **C)** Band integration from the SDS-page and **D)** data from the fluorescence plate reader were used to calculate aggregation temperatures: HORMAD2-uvGFP (SDS-PAGE: 68.89, Plate Reader: 70.03), GFP (SDS-PAGE: 78.11, Plate Reader: 78.16). E) HORMAD2-uvGFP was combined with 1 mM $MgCl_2$, 5 µM ZnAc, 100 µM $MnCl_2$, 90 ng/uL genomic Bovine dsDNA restriction enzyme digested into short dsDNA, DNA (as before) combined with metal compounds (as before), or 5 mM ATP and GFP-BASTA was performed again at 69°C and supernatants were analysed on a plate reader. RFU, Relative fluorescence units; GFP, Green fluorescent protein

4.2.14 DSF-GTP

Differential scanning fluorometry of a GFP-tagged protein (DSF-GTP) was also intended to be used to investigate small-molecule interactions with HORMAD2-uvGFP. Recombinant protein is heated over a large temperature gradient under constant monitoring for GFP fluorescence. Unfolding of the GFP-tagged protein causes a change in GFP fluorescence, creating a peak, followed by a large peak at 81°C coinciding with the denaturing of the GFP tag itself. Under normal conditions in phosphate buffer, HORMAD2 produced no peak in the assay, however the GFP peak appeared to be slightly left-skewed, indicating the HORMAD2 peak could be lost within the GFP peak. To try to reveal the HORMAD2 peak, different buffers were trialled and various metal compounds were added, but no peak was produced.

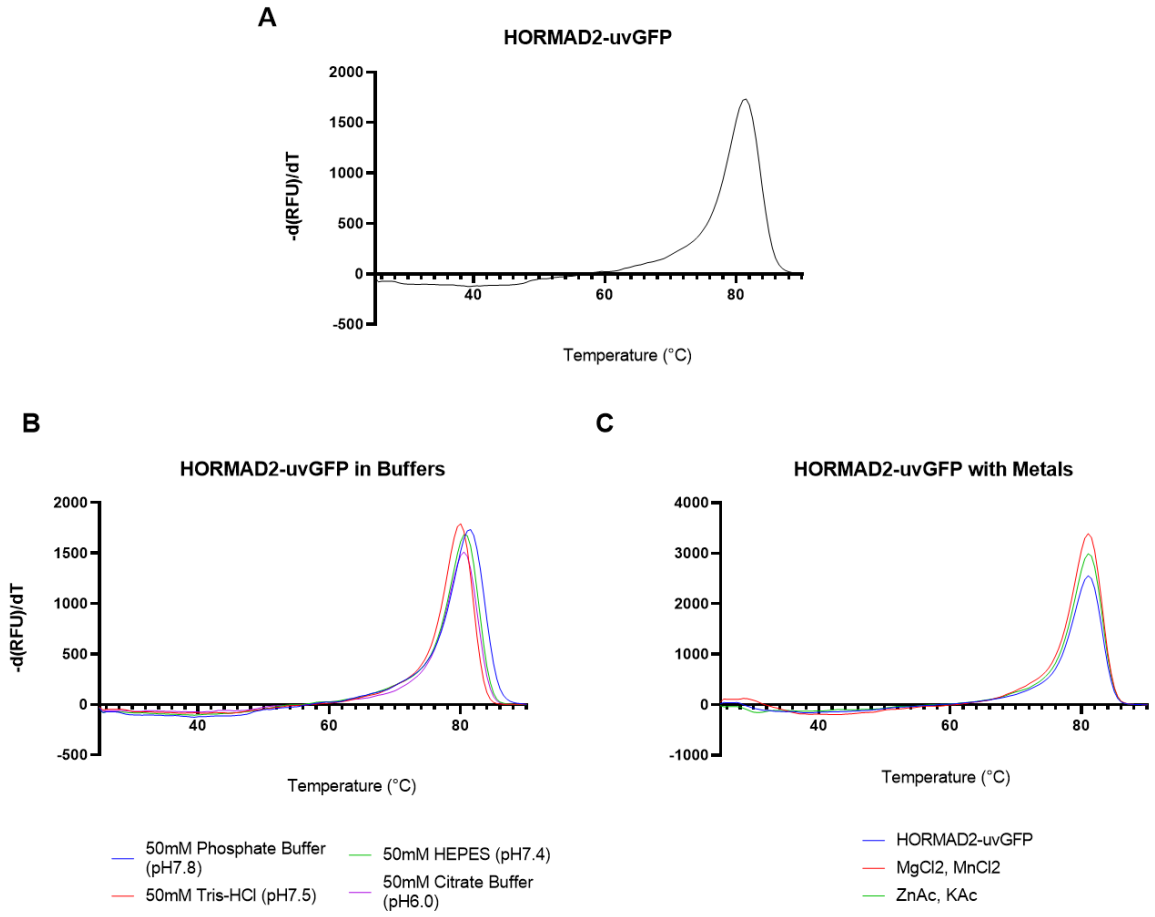

Figure 4.16: DSF-GTP of HORMAD2 to show molecular interactions. HORMAD2-uvGFP was heated from 25-90°C with constant monitoring of GFP-fluorescence. **A)** The protein in standard 50mM phosphate buffer (pH 7.8) did not produce a peak for HORMAD2. **B**) Trialling different buffers: 50 mM phosphate buffer (pH 7.8), 50 mM Tris-HCl (pH 7.5), 50 mM HEPES (pH 7.4) 50 mM Citrate Buffer (pH 6.0) did not produce a peak. **C)** Addition of 100 µM MgCl and 10 µM MnCl or 100 µM KAc and 10 µM ZnAc in 50 mM HEPES (pH 7.5) did not produce a peak.

4.3 Discussion

HORMAD2 is a poorly characterised meiotic protein which may play a role in HCC. Preliminary *in vivo* data has suggested that HORMAD2 expression is altered in association with changes in tumour size in response to a high fructose diet, and *in silico* data suggests that HORMAD2 is tumour-suppressive in human HCC. Based on its meiotic function, HORMAD2 was hypothesised to influence DNA repair and/or cell proliferation.

Manipulation of HORMAD2 Expression in HCC

To investigate HORMAD2 function *in vitro*, HCC cell lines were modified to overexpress or knockdown HORMAD2 expression. Although several attempts were made using a variety of techniques, constitutive overexpression of HORMAD2 consistently failed to produce viable cells (**Figure 4.2**). Furthermore, unpublished observations (data not shown) indicated that inducible overexpression of HORMAD2 initially caused slower growth of cells during the generation of stable cell lines. However, once the cell lines were established and ready for experimentation, the effect was quickly lost. Once successful, inducible overexpression of HORMAD2 in HCC cell lines was very weak and caused a barely-appreciable increase in HORMAD2 expression (**Figure 4.3**). Similarly, knockdown of HORMAD2 seemed to promote more rapid growth during stable cell line generation, but the effect was lost shortly after.

These observations, though unsupported by experimental data, are supportive of *in silico* findings, which suggest that HORMAD2 could be cell-cycle repressive based on correlation analyses **(Chapter 3.2.6)**. Based on the observations of HORMAD2 overexpression, it is likely that, if this is the case, negative selective pressure of HORMAD2 expression led to survival of only those cells that produced weak HORMAD2 overexpression. However, given that a similar effect was observed in HORMAD2 knockdown, another possibility is that the HCC cell lines were somehow able to adapt to changes in HORMAD2 expression. In either case, an effective method of transient overexpression and knockdown of HORMAD2 may be required to observe these effects. Unfortunately, due to difficulties with transfection into HCC cell lines, this was not achieved in this project.

Fructose and HORMAD2

Increased fructose consumption in the "western diet" has been linked with the increasing prevalence of HCC due to its associated toxic effects on the liver such as increased fatty acid production, oxidative stress and insulin resistance (18). Indeed, replacing the carbohydrate content of cell culture media with fructose is toxic to HCC cell lines. In this study, it was found that fructose also caused down-regulation of HORMAD2 (**Figure 4.1**). This is consistent with preliminary data which shows that HORMAD2 is down-regulated *in vivo* in fructose-fed, adiponectin-knockout mice (**Figure 1.1B**). It is possible that this effect associated with fructose was not causative, and instead the toxicity of fructose indirectly

caused down-regulation of HORMAD2 as a result of cellular shut-down. However, it is also possible that fructose is able to down-regulate HORMAD2 via direct effects on cellular pathways. For example, fructose has been shown to attenuate MAP-kinase, JNK and insulin signalling pathways *in vitro* in primary hepatocytes (233). If HORMAD2 is indeed tumour-suppressive, and responds to oxidative stress as *in silico* data suggests, this could in part explain some of the association of fructose consumption and increased risk of HCC.

To determine whether HORMAD2 expression could have a reciprocal effect on HCC cell sensitivity to fructose, an MTT assay was used to assess cell viability in HORMAD2-overexpressing and knockdown cells grown in fructose-media **(Figure 4.8)**. No significant effect of HORMAD2 expression was observed. Thus, under the conditions tested, changes in HORMAD2 expression alone were not sufficient to affect fructose-sensitivity.

Nuclear Localisation

Nuclear fractionation of HCC cell lines expressing HOMRAD2-6HIS indicated that HORMAD2 localises primarily in the nucleus. This is consistent with nuclear localisation demonstrated in the testes (32, 36), and with the C-terminal nuclear localisation signal detected in an *in silico* characterisation of HORMAD2 **(Chapter 3.2.11)**. Considerable antibody staining for HORMAD2-6HIS was also present in the cytoplasmic fraction which has not been shown previously. Immunofluorescence staining of cells was attempted to support these findings, however staining for the 6HIS tag produced too much background (data not shown).

Cytoplasmic staining of HORMAD2 could be due to protein which has not yet been transported to the nucleus, or it could suggest alternative cytoplasmic functions of the protein. It is not unheard of for proteins to have diverse functions which involve translocation between cellular compartments, such as PFKFB3 a glycolytic enzyme which can also translocate to the nucleus to affect the expression of DNA repair enzymes (110). Further, the C-terminus of HORMAD2 which contains its nuclear-localisation signal also contains its HORMA closure motif, and is also in close proximity to an S/T-Q site for phosphorylation by PIKK-family kinases. Thus, it is possible that protein-protein interactions or post-translational modifications could sterically regulate nuclear localisation of HORMAD2.

Cell Proliferation

When studied using the knockdown and inducible overexpression systems, HORMAD2 expression had no significant effect on cell proliferation in 2-dimensional cell culture. This was particularly unexpected, as the failure of the constitutive-construct to produce viable cells suggested that the cell cycle was being directly impacted by HORMAD2 overexpression. Indeed, physical observations during cell culture initially indicated that inducible overexpression of HORMAD2 led to slower growth rates,

while HORMAD2-knockdown produced faster-growing cells (data not shown). However, these observations were not supported by any experimental data.

Tumoursphere assays were also performed, which involves culturing cells under conditions that prevent them from adhering to the plastic surface of the cell culture plate, and thus promotes 3-dimensional spheroid growth. In the tumoursphere assays of unirradiated HORMAD2-knockdown cell lines, HORMAD2-knockdown again did not seem to significantly affect the number of spheres formed. However, it was clear visually that HORMAD2-knockdown promoted the production of much larger and more aggregated spheres (**Figure 4.9B**). Spheroid aggregation is a common problem in 3-dimensional cell culture and is notorious for making spheroid counting difficult.

However, little research has been performed to investigate the factors that promote spheroid aggregation, and whether this is indicative of increased metastatic potential, for example. Two recent studies have demonstrated that protein kinase N1 (PKN1) and -N2 (PKN2) deletion affects spheroid compaction of embryonic fibroblasts in suspension culture through down-regulation of N-cadherin and integrins α5 and αV (234, 235). Thus, it is not unreasonable to suggest that HORMAD2 could influence the expression of cellular adhesion proteins in a similar way to suppress tumourigenesis.

This is supportive of the initial unpublished observations described above and of the *in silico* data, however it is unknown why these findings didn't translate into the cell proliferation assays performed. One possible explanation is the specific culturing conditions for the tumoursphere assay, specifically the addition of a number of hormones and growth factors, namely: hEGF, bFGF, insulin, and the B27 Supplement for neuronal growth. It is possible that HORMAD2 acts specifically downstream of one of these growth factors or hormones, and thus their inclusion in the cell culture media triggered the effect on cell growth. Alternatively, it could be the 3-dimensional nature of growth in the assay. 2-dimensional cell culture techniques are known to be inferior in their ability to recreate the complexity of *in vivo* cell-cell and cell-ECM interactions, and alter the cell division processes and gene expression (236). Thus, 3-dimensional cell culture techniques may be better able to show the effect of HORMAD2 expression on cell growth.

Another possibility, which the Hebbard lab is now investigating, is that HORMAD2 could regulate cancer stem cell number. The tumoursphere assay is used to detect cancer stem cells, because only these can survive and proliferate in environments where they are unable to adhere (237). Indeed, HORMAD2 is known to function alongside BRCA1 in meiosis (36), and BRCA1 has been demonstrated numerous times to be a negative-regulator of cancer stemness (238, 239). Thus, HORMAD2 may function alongside BRCA1 to negatively regulate CSC number, which would explain the results of the tumoursphere assay.

HORMAD2 Expression and DNA Damage

HORMAD2 is known to interact directly or indirectly with a number of DNA-damage response proteins including key members of the double-stranded break response, ATR and BRCA1, in its meiotic functions. Thus, it was hypothesised that if HORMAD2 plays a role in HCC, it is likely to be in the DNA damage response. This hypothesis was tested in a number of ways.

The viability of HCC cells was analysed using MTT cell viability assays, colony formation assays and tumoursphere assays. In MTT and colony formation assays, HORMAD2 was shown to have no effect on cell viability after treatment with Cisplatin, Sorafenib, H_2O_2, or radiation. Tumoursphere assays also showed no significant change in tumoursphere number after radiation, however the larger and more aggregative tumourspheres discussed earlier persisted throughout 1.8 Gy and 5.4 Gy radiation treatment. Thus, knockdown of HOMRAD2 allows larger aggregative spheroids to persist after low and medium dose radiation, but it does not seem to influence resistance to chemotherapeutics, at least in 2-dimensional culture.

Next, comet assays were used to assess the overall amount of DNA damage remaining in each cell line 1 hour after treatment of radiation of hydrogen peroxide. In both cases, HORMAD2 knockdown was associated with significantly decreased levels of DNA damage. It remains unclear whether HORMAD2 overexpression has a similar effect due to high variability of the control cell line. This may indicate that the doxycycline added to the PCW57.1 cell lines affected comet appearance in the assay.

Finally, western blots were used to evaluate activation of DDR pathways in response to radiation. Phosphorylation of ATM, ATR and P53 were evaluated approximately 1hr after radiation. No appreciable difference was observed between the cell lines. It is possible that HORMAD2 does affect activation of these pathways but over a much faster timescale after damage has occurred, however due to limitations of transporting cells after radiation, this could not be investigated further for this project.

Overall, the results here somewhat suggest that HORMAD2-knockdown promotes the repair of radiation and H_2O_2-induced DNA damage. However, this change in DNA repair is not sufficient to affect cell viability, and HORMAD2 was not found to affect signalling within the ATM, ATR or p53 pathways. One possible explanation for these findings is that HORMAD2 activity acts on a very short timescale to attenuate the rate of DNA repair, without affecting the overall capacity for DNA repair. This would mean that changes in pathway activation would need to be analysed much more rapidly, and a comet assay over multiple timepoints could show that, over a longer timescale, overall DNA-damage equalises and thus cell-viability over 24 and 48 hours remains unaffected. In any case, these conclusions require further experimentation to support.

One of HORMAD2's suggested roles in meiosis is inter-homolog bias (25, 32), which may provide insight into a mechanism of DNA-repair attenuation. A key difference between the repair of double-

stranded breaks in somatic cells and meiocytes is that in somatic cells, the genetically identical sister chromatid is the ideal partner for template-directed repair, while in meiocytes, repair with the sister chromatid is inhibited to create a bias towards the homologous chromosome and thus facilitate crossover events (37-39). Thus, if HORMAD2's function in meiosis were to repress inter-sister recombination, this could translate to an overall repression of double-stranded break repair in somatic cells. It remains unclear how this would be achieved mechanistically however, and further studies are required to validate this hypothesis.

Recombinant HORMAD2-GFP Expression and Assays

To further characterise HOMRAD2 function, recombinant protein was expressed in *E. coli*. It is clear that the prokaryotic *E. coli* protein expression system is not ideal for the generation of HORMAD2 recombinant protein for functional analysis. The yield of correctly-folded protein was very small, which is indicative of the prokaryotic system being unable to reproduce the complexity of eukaryotic protein folding. Additionally, the disordered C-terminal tail of HORMAD2 made it prone to proteolysis (240, 241), and further reduced the yield of usable protein. Finally, the use of this recombinant protein for functional analysis was further limited by the fact that functions of HORMAD2 are likely to require eukaryotic systems of post-translational modification such as phosphorylation (36), and the presence of a bulky GFP-fusion tag has the potential to impede molecular interactions.

Nonetheless, a sufficient yield of seemingly correctly-folded protein was acquired after optimisation of the expression and purification process. A GFP-EMSA supported DNA-binding of recombinant HORMAD2 to a variety of DNA samples. Band shifts were most obvious in the DNA, but less-so in the protein fluorescence, which can likely be attributed to the protein being a mixture of correctly and non-correctly folded HORMAD2. Non-specific DNA binding was supported by evidence of band-smearing when combined with a BioO short dsDNA, bovine genomic DNA, and human genomic DNA.

GFP-BASTA was also used to characterise HORMAD2 based on its aggregation temperature (T_{agg}). The T_{agg} of a protein is closely-related to its melting point (T_m), and the T_{agg} of HORMAD2 was calculated to be 69°C. This is supported by a bioinformatic, sequence-based prediction of 66.7°C. It is possible that this melting point reflected denaturing of the GFP-fusion tag, which usually unfolds at approximately 81°C, however the DSF-GTP results indicate that the unfolding temperature of GFP was unaffected by the fusion of HORMAD2. GFP-BASTA was then performed in combination with various common co-factors of DNA-binding proteins, as well as DNA itself and ATP. The expected result of this is that a shift in the T_{agg} of a protein is indicative of a stabilising or destabilising interaction with another molecule, however, none of these molecules produced a shift in T_{agg} of HORMAD2. Thus, this assay failed to produce evidence of any molecular binding partners of HORMAD2. However, the evidence that the prokaryotic system produced a low yield of correctly folded protein means that the recombinant HORMAD2-GFP used in this assay was likely contaminated by large amounts of non-

functional protein, which would have added a great deal of noise to the experiment, and could explain why no shift in T_{agg} was observed in GFP-BASTA.

Finally, a DSF-GTP assay was also used to characterise HORMAD2, though unsuccessfully. The DSF-GTP assay essentially involves gradient heating of a protein fused to GFP. At the temperature where a protein unfolds, the GFP-fluorescence is usually affected, which results in a peak on the derivative output, followed by a large peak at 81°C indicating unfolding of GFP itself **(Figure 4.16)**. The results of this project failed to show a peak for HORMAD2, which is likely due to either the relatively high thermal stability of HORMAD2 meaning that its peak is lost within the much larger GFP peak, or because the disordered C-terminal of HORMAD2, to which GFP is fused, prevented any attenuation of GFP fluorescence during HORMAD2 denaturation. For this reason, various buffers were used to try to separate the two protein peaks, but to no avail.

The results of these molecular assays are likely to have been hindered by a number of factors. The evidence that the prokaryotic system produced a low yield of correctly folded protein means that the recombinant HORMAD2-GFP used in these assays was likely contaminated by large amounts of non-functional protein, which would have added a great deal of noise to each experiment. This could explain why no shift in T_{agg} was observed in GFP-BASTA.

4.4 Conclusion

HORMAD2 was hypothesised to act as a tumour-suppressor to HCC based on *in silico* data. While many of the techniques used in this project were incompatible with HORMAD2, the results acquired herein are supportive of HORMAD2 being a tumour-suppressor. It was found that knockdown of HOMRAD2 promotes the formation of large, aggregative spheroids in 3-dimensional culture, which could indicate increased metastatic potential of these HCC cells. Evidence also suggests that HORMAD2 suppresses DNA repair. This suggests a mechanism of tumour suppression whereby HORMAD2 encourages hepatocytes to “shut down” in response to DNA damage rather than repair and survive. Furthermore, fructose appears to down-regulate HORMAD2, which indicates that this protein plays a role in fructose-promoted HCC. Overall, much further investigation is required to validate these hypotheses, but HORMAD2 does indeed appear to influence HCC.

Chapter 5: Molecular Pathways of Smoke-Induced OPC Pathogenesis

5.1 Preamble

The addition of this chapter of the thesis was catalysed by a collaboration with Dr. Kylie Lopes Floro who required assistance in the bioinformatic analysis of sequencing data for an oral-pharyngeal cancer project. Thus, I was not involved in the planning or execution of this project, and all work prior to the bioinformatic analyses were conducted by Dr. Kylie Lopes Floro and A/Prof. Lionel Hebbard as outlined in **Section 5.3.**

5.2 Introduction

Clinically, patients who continue to smoke during radiation therapy for OPC are known to suffer worse outcome and have increased incidence of metastasis (171). These effects are believed to be due to *in vivo* effects of increased hypoxia in cigarette smokers, which decreases the efficacy of radiation treatment. However, *in vitro* models suggest that both smoke and radiation can independently induce the formation of more invasive and/or more stem-cell like phenotypes (172, 174, 189). Thus, the goal of this investigation was to elucidate molecular pathways involved in the more aggressive cancer phenotype of patients who continue to smoke during radiation treatment, and whether the smoke and radiation can act synergistically to enhance tumour invasion and stemness.

5.3 Methods

The cell culture, experimental treatments, and RNA collection for this chapter were performed by Dr. Kylie Lopes Floro, Dr. Miriam Wankell and A/Prof. Lionel Hebbard. For context, the methods of cell culture and experimental conditions are outlined below.

FaDu OPC cells were cultured in DMEM-F12 cell culture media, with 10% FCS and 1% Pen/Strep. Cells were split into 12 flasks to create 4 treatment groups in triplicate: 1) Untreated control, 2) Smoke-treated, 3) Radiation-treated, 4) Smoke and radiation-treated.

Cigarette smoke-treated media was prepared as follows; cigarettes (one pack, 25 cigarettes, of JPS red label) were burnt to within 2cm of the butt, and the smoke, and thus tar/nicotine, were bubbled through 500mls of DMEM-F12 under negative pressure. The DMEM-F12 was subsequently filtered prior to use. Cells were then cultured using this smoke-treated DMEM-F12 in alternating dilutions of 1/50 then 1/25, changing every 2 days for the duration of the experiment. This was done because a previous MTT assay had shown that both of these concentrations could reduce, but not completely prevent, proliferation of the FaDu cell line.

Cells were irradiated at 1.8 Gy per fraction in flasks using 6mV photons using an Elekta linear accelerator. The predicted and delivered dose were compared using Gafchromic EBT3 film (Alpha XRT); the difference was found to be within the uncertainty of the film measurement (± 2 %). The average discrepancy with the planning system is -0.14 %. Cells received a total of 27Gy over three weeks. Cells were allowed to recover for three days prior to the collection of RNA (Bioline Isolate II RNA Mini Kit, BIO-52072).

5.3.1 Principal Component Analysis

In total, the gene expression was quantified for 21 979 genes across 12 samples. Principal component analysis (PCA) of the gene expression data showed that Control, Radiation, Smoke, and Smoke and Radiation all formed distinct groups, and collected tightly within each group **(Figure 5.1)**. This indicates that within each triplicate treatment, there was very little variation in gene expression, but there were large differences between samples. Additionally, as expected, cells treated with both smoke and radiation showed the most variation from control cells.

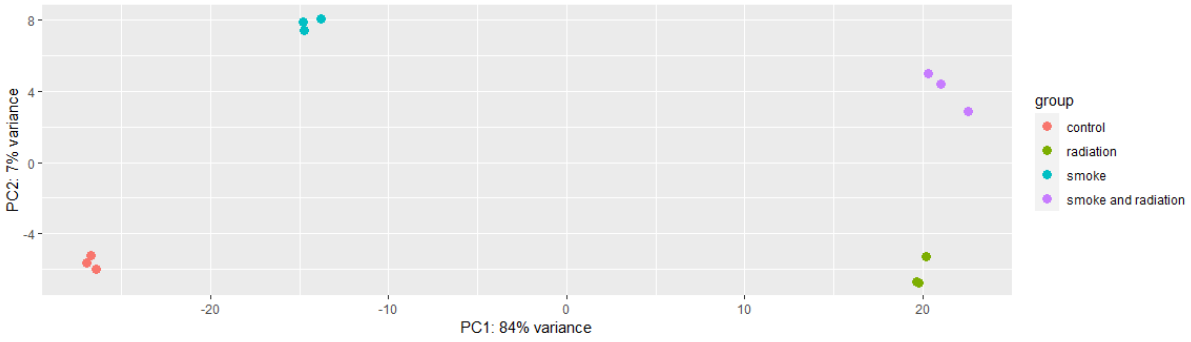

Figure 5.1: Principal component analysis plot of FaDu cell gene expression. Cells were either untreated (control), or treated with 27 Gy radiation (radiation), cigarette-smoke treated media (smoke), or both (smoke and radiation). The plot was generated in R using variance-stabilised, log-transformed data from DESeq2.

5.3.2 Differential Expression Analysis

Differential expression was analysed between the treatment groups as pairwise comparisons for changes of greater than 25% in gene-expression to filter the results. Treatment with radiation caused changes in a far greater number of genes than treatment with cigarette smoke-treated media. Treatment with both radiation and smoke caused the greatest number of differentially expressed genes when compared with control cells **(Figure 5.2)**.

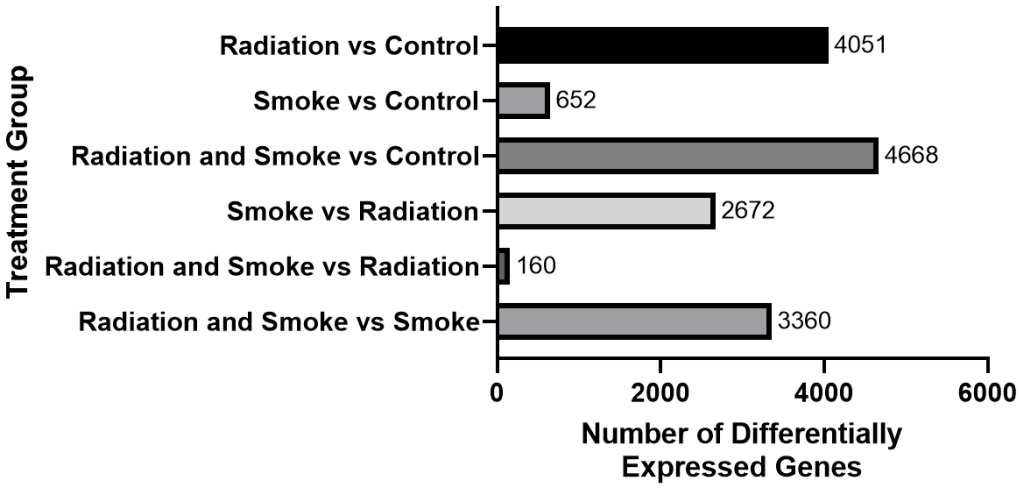

Figure 5.2: Number of differentially expressed genes between treatment groups. Differential expression was analysed using DESeq2. Significant values were identified as $p < 0.05$.

A Venn diagram was generated of differentially expressed genes in the Smoke, Radiation and Smoke and Radiation groups relative to control **(Figure 5.3)**. 65.42% of the genes differentially expressed in cells that received the combination treatment also occurred in the radiation-only treatment, while only 10.22% of genes were affected by smoke. However, a large fraction of differentially-expressed genes (21.26%) were unique to the combination treatment.

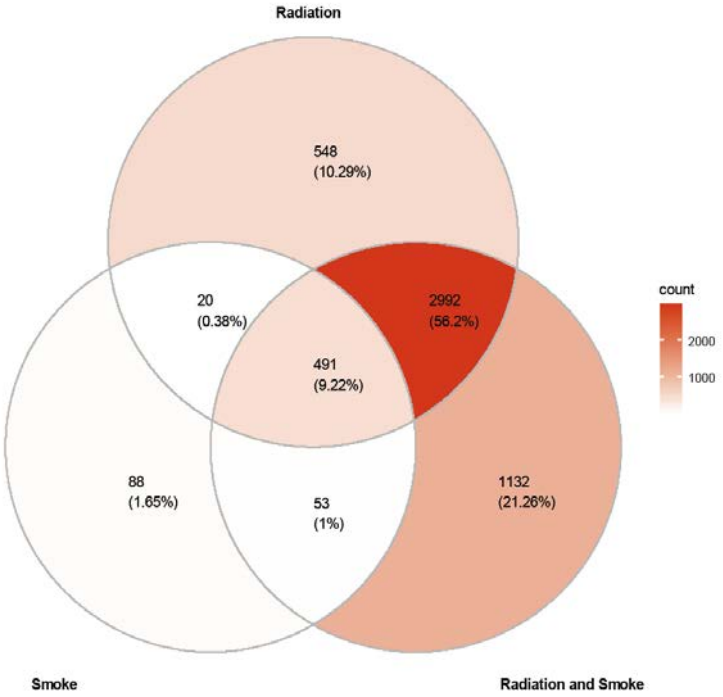

Figure 5.3 Venn diagram of differentially expressed genes in treatment groups versus control. Significantly differentially expression genes ($p < 0.05$) were extracted using DESeq2 for the radiated cells, cigarette smoke-treated cells, and the combination treatment of radiation and cigarette smoke in all in comparison to untreated control cells.

In order to visualise the vast amount of differentially-expressed genes, multiple steps of enrichment analysis were performed. For each pairwise comparison, significantly-differentially expressed genes were analysed for Gene Ontology (GO) biological process (BP) terms which appear densely at the top of a ranked list using the **G**ene **O**ntology en**RI**chment ana**L**ysis and visualization (GORILLA) tool. These enriched terms were then further filtered for redundancy using **RE**duce and **VI**sualise **G**ene **O**ntoloty (REVIGO). Finally, select GO-BP terms were then chosen to display in **Figure 5.4**, however the entire list can be found in **Appendix D-I**. Ranking was performed using the test-statistic of the DESeq2 Wald test, which accounts for both significance (p value) and the direction of change (up vs down-regulated), and thus, each analyses was performed in both directions to evaluate up and down-regulated pathways.

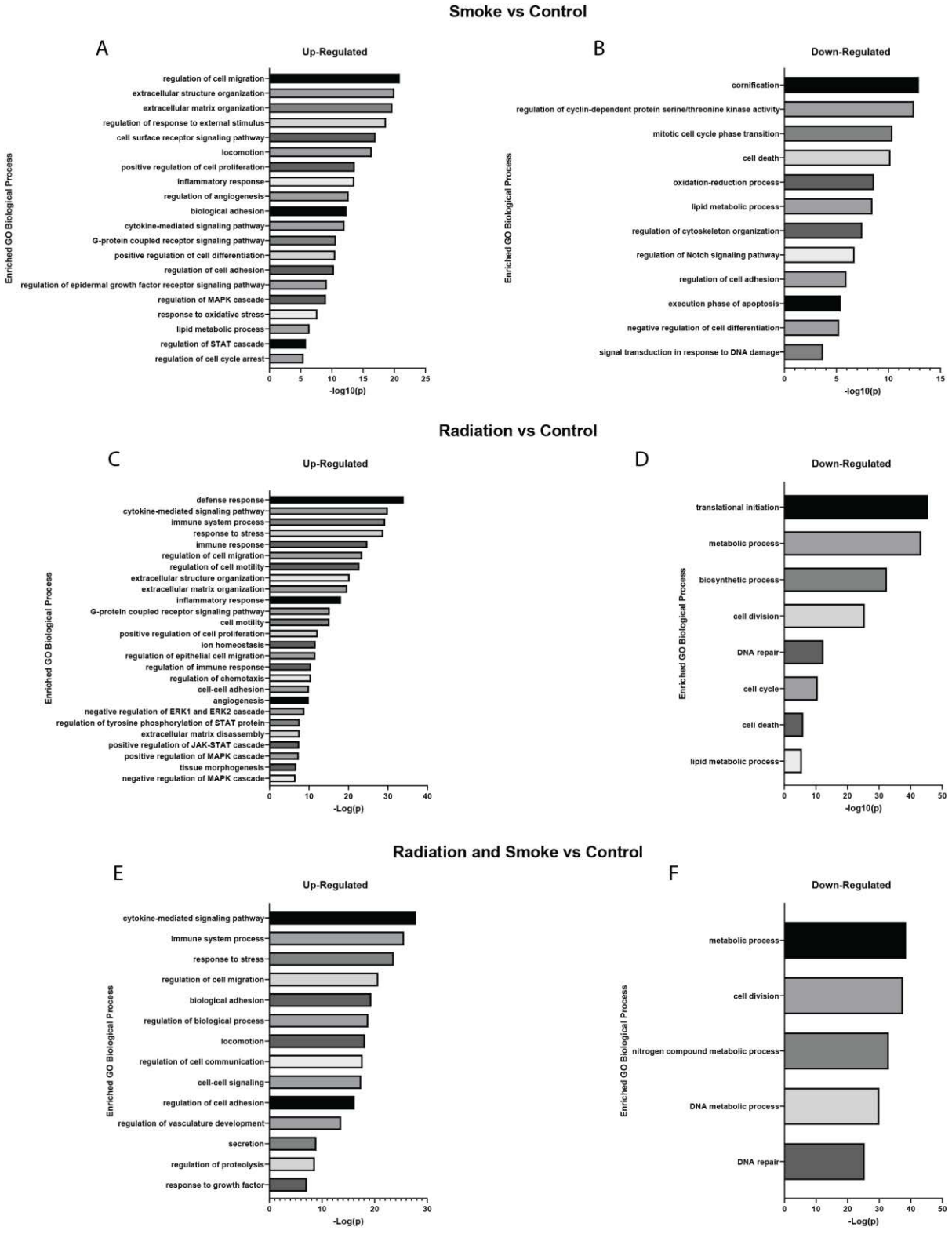

Figure 5.4: Pathway enrichment: Radiation and smoke treatment compared with untreated cells. Differential expression was performed on cells treated with **A and B)** cigarette-smoke treated media, **C and D)** 27 Gy radiation, or **E and F)** both smoke and radiation, all compared with untreated controls using DESeq2. Pathway enrichment was performed using GORILLA.

Firstly, each treatment group was compared with the untreated cells for up or down-regulated pathways as described above. All three treatments promoted the expression of genes involved in pathways of cell motility, migration, and/or locomotion, as well as cell-cell adhesion, and angiogenesis. Radiation and smoke combination treatment also promoted secretion and proteolysis. Down-regulated pathways in all treatments were related to cell division and cell death. Interestingly, smoke treatment also down-regulated genes involved in cellular adhesion **(Figure 5.4)**.

Next, further investigation was made into the combination treatment of smoke and radiation together. This was done in two ways. Firstly, pairwise differential expression was performed between cells that received the combination treatment compared with those that received only radiation, and pathway enrichment was performed as before. Additionally, the 1132 genes that were uniquely differentially expressed in the combination treatment group **(Figure 5.3)** were separated into up and down-regulated genes and analysed for pathway enrichment in an un-ranked analysis with the entire set of analysed genes used as background.

Comparing combination treatment with radiation alone, the combination treatment caused cells to have higher metabolic activity, and increased cell-signalling and communication. The combination treatment also caused enhanced up-regulation of cell adhesion, migration, cytoskeleton organisation and angiogenesis when compared with radiation treatment alone. Down-regulated pathways were also related to cell adhesion and extracellular structure organisation. Some of the most highly up-regulated genes included: ALDH1A3 (p = 2.24E-40) and ALDH3A1 (p = 1.67E-34), SLPI (p = 6.42E-09) and MMP9 (p = 1.44E-08). Strongly down-regulated genes included: IGFL1 (p = 7.37E-63), IFI6 (p = 4.51E-11), and STAT1 (p = 1.83E-04) **(Figure 5.5)**.

Some genes were uniquely altered only after combination treatment of smoke and radiation, and were not detected as significantly differentially expressed relative to control cells after smoke or radiation treatment alone. This was visualised in **Figure 5.3.** Up-regulated genetic pathways included metabolic pathways, and interestingly embryonic-placental development-related cell differentiation, specifically from up-regulation of SNAI1 (p = 0.016) and SOX15 (p = 7.43E-06). Some of the most highly up-regulated genes in this group included cathepsin A (CTSA) (p = 2.78E-26), claudin 4 (CLDN4) (p = 3.28E-22) and secretory leukocyte protease inhibitor (SLPI) (p = 1.27E-21). Cells that received combination treatment also showed unique down-regulation of DNA-repair, cell cycle and metabolic pathways, including specific genes such as DNA-PKcs (PRKDC) (p = 1.48E-11), Transformation/transcription domain-associated protein (TRRAP) (p = 2.31E-14) and DNA polymerase epsilon catalytic subunit (POLE) (p = 5.79E-13) **(Figure 5.5)**.

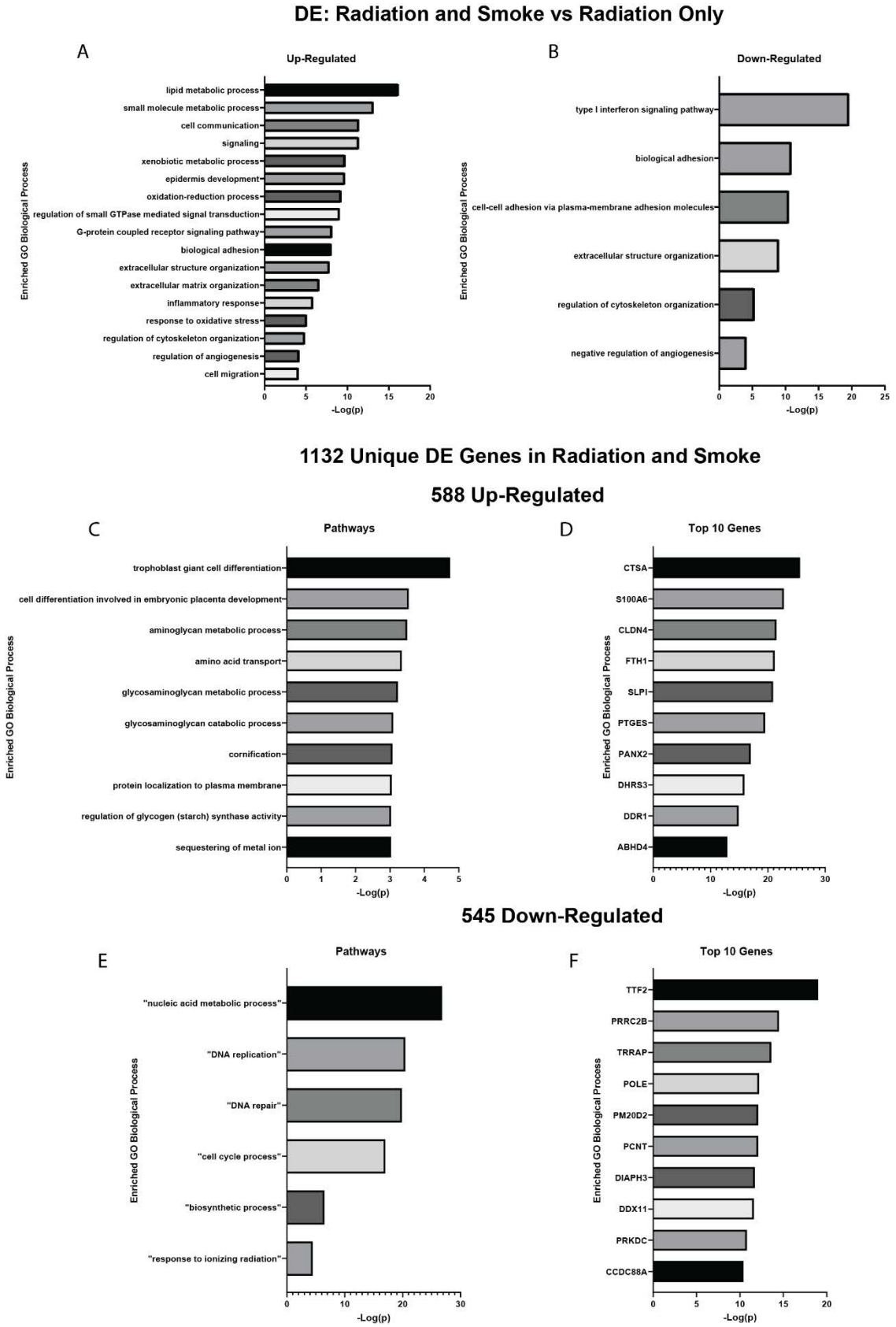

Figure 5.5: Uniquely altered pathways in cells treated with both radiation and smoke. Differential expression was performed on: **A and B)** cells treated with both radiation and smoke compared to those treated with radiation alone, **C-F)** cells treated with both smoke and radiation compared to untreated cells, filtered to remove differentially expressed genes after smoke or radiation treatment alone. Differential expression analysis was performed using DESeq2, and pathway enrichment was performed using GORILLA.

Next it was necessary to understand how the known pathways of metastasis were affected by each treatment. For this next analysis, genes were hand-selected based on a number of publications which have attempted to identify the key markers of invasion, stemness, endothelial-mesenchymal transition (EMT), growth and survival, and angiogenesis pathways (242-246). Heatmaps were generated based on expression data for each of the genes in these curated pathways **(Figure 5.6)**.

Effects on markers of invasion were variable **(Figure 5.6A)**. Unexpectedly, motility-related proteins RHOA and RAC1 were both down-regulated in response to radiation and the combination treatment, which is in contrast to the increased migration of these cells *in vitro*. However, oncogenic cellular adhesion molecules, particularly CDH11 and ITGA5, were strongly up-regulated in all cells that received radiation treatment, and ITGA5 expression was increased synergistically after combination treatment. SDC1 was also increased but only after combination treatment. A similar trend was found for most MMPs, where MMP2, MMP9 and MMP14 again showed synergistic increase in expression after combination treatment. Cathepsin D was mostly unaffected by smoke or radiation treatment alone, but in combination these treatments caused relatively strong up-regulation.

Markers of cell stemness showed mostly small and variable changes **(Figure 5.6B)**. CD44 expression was higher in radiated cells, but slightly higher again in cells that underwent combination treatment, while other cell-surface makers were mostly unaltered. The effects of the treatments on the various ALDH molecules were also variable, however combination treatment synergistically lowered ALDH18A1, ALDH1A1, and ALDH1B1 expression, and synergistically increased ALDH2. Interestingly, smoke treatment alone strongly increased ALDH3A1 expression, while radiation reduced its expression relative to control cells, and the combination treatment lowered the reduction observed in radiated cells.

Little or no synergistic effect was observed on genes within the curated EMT pathway **(Figure 5.6C)**. However, notably, cigarette smoke caused down-regulation of KRT5, and this down-regulation was much stronger in cells treated with radiation or combination treatment. Similarly, smoke treatment caused up-regulation of ETS1, which was stronger in radiation and combination-treatment cells.

Angiogenesis-related genes showed strong changes in response to treatment **(Figure 5.6D)**. VEGFA/B/C were all increased in radiated cells, and all appeared to show synergistic increase which was strongest for VEGFA. VEGFD remained unaffected, while PDGFA was down-regulated by all treatments.

Finally, of the genes investigated relating to growth and survival, there was mostly no changes **(Figure 5.6E)**. However, smoke and radiation both caused up-regulation of CDKN1A, and this effect was synergistic in combination treatment.

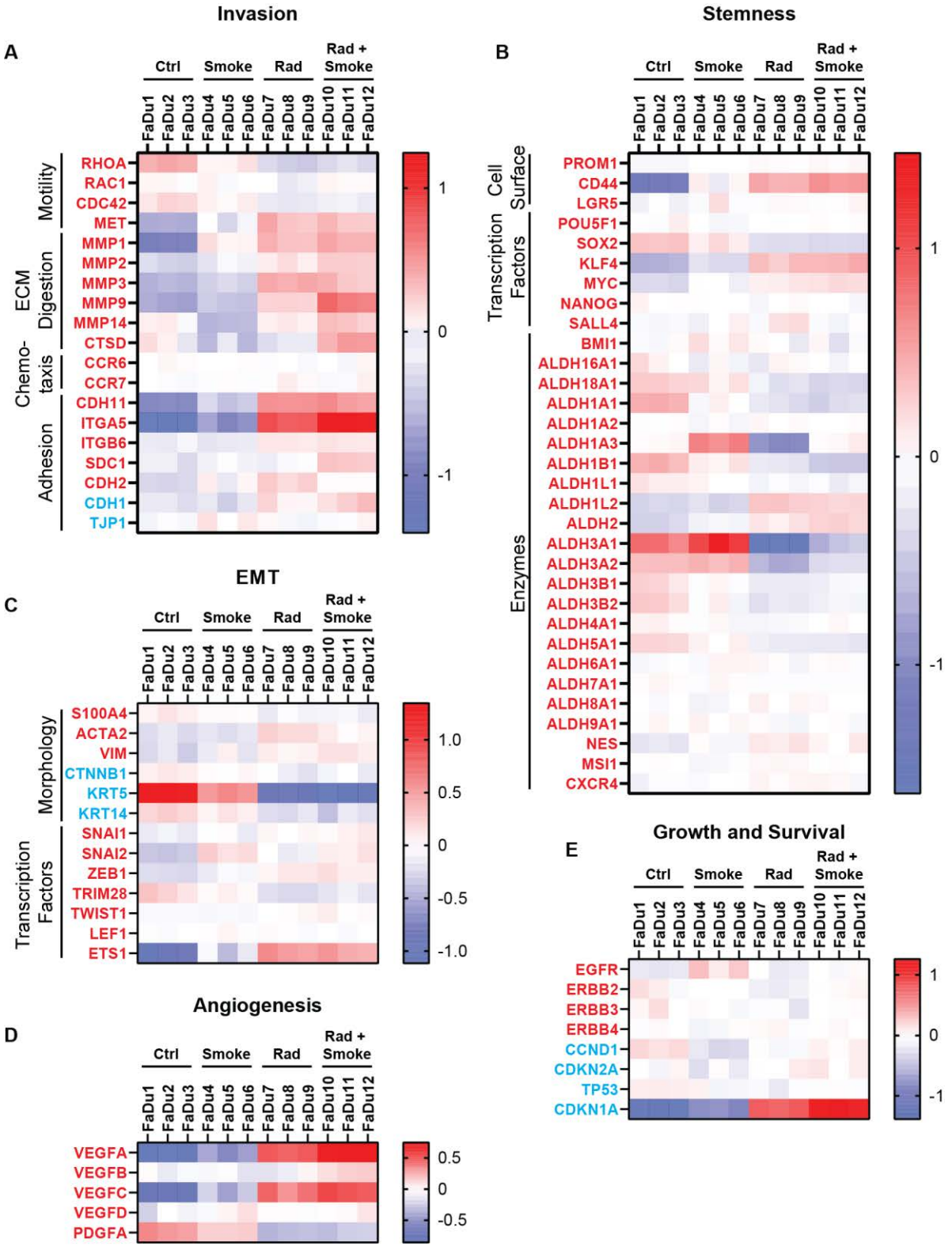

Figure 5.6: Heatmap: Gene expression in curated pathways of tumour invasion. Pathways are separated into known markers of **A)** invasion, **B)** stemness, **C)** epithelial-mesenchymal transition (EMT), **D)** angiogenesis, and, **E)** growth and survival. Red gene names indicate oncogenes, while blue genes are tumour suppressive, or otherwise usually attenuated in cancer. Values are log2 fold-changes from the mean and were normalised with variance stabilising transformation, as explained in chapter 2.1.3.5. Ctrl, Control; Rad, Radiation; EMT, Epithelial-mesenchymal transition

5.4 Discussion

Patients who continue to smoke during radiation therapy are known to have decreased loco-regional control of their tumours after radiation treatment. It has been assumed that the mechanisms of this are related to increased hypoxia, which decreases radiation therapy efficacy and promotes angiogenesis (171). However, *in vitro* observations in cell culture, which are incapable of reproducing the complex hypoxic environment of solid tumours, have demonstrated that radiation and cigarette smoke can independently promote stem-cell characteristics in OPC cells (189). Thus, the purpose of this investigation was to evaluate whether radiation and smoke promote more invasive and stem-like OPC phenotypes *in vitro*, and whether their effects are synergistic.

In vitro Experiments

In vitro experiments were performed by Dr. Kylie Lopes Floro to investigate OPC cell migration and invasion following the smoke, radiation, and combination treatment described for this experiment. The unpublished results of these experiments show that both treatments independently and in combination created more invasive and migratory cells, with a mesenchymal morphology. This next-generation sequencing data of those cells supports an increase in migration, however it does not appear to be acting through canonical motility pathways **(Figure 5.6)**, but rather the changes in migration seem to be related mostly to changes in cellular adhesion such as ITGA5 over-expression, as well as increased secretory protease activity and cytokine-mediate chemotaxis. There is also strong *in silico* support for the observations of EMT, particularly from increases in Vimentin expression and actin α2.

Cigarette Smoke Treatment

Compared with radiation alone and the combination treatment of smoke and radiation together, cigarette smoke alone caused the smallest number of differentially expressed genes. Interestingly, of the genes that were significantly differentially expressed in smoke-treated cells, 85% were also differentially expressed after radiation treatment alone. This suggests that both cigarette smoke and radiation affect cellular processes similarly.

Cigarette smoke treatment caused up-regulation of genes involved in migration and cell adhesion. The key genes responsible for this enrichment were mostly metalloproteinases (MMP1, MMP10, ADAM8, ADAM12) and integrins (ITGA2, ITGA5). Interestingly, cigarette smoke caused very strong up-regulation of not only CD44 ($p = 6.13E-47$) but also the hyaluronan synthases 2 and 3 (HAS2, HAS3) ($p = 1.81E-12$ and $8.63E-62$ respectively) **(Appendix D)**. CD44 interaction with its primary ligand, hyaluronic acid promotes cell migration and proliferation via Ras, MAPK and PI3K signalling. Additionally, CD44 has a known role in maintaining stemness of cancer stem cells (187).

It is unknown mechanistically how cigarette smoke can induce these changes. However, one important component of cigarette smoke is nicotine, which has been shown to induce invasive phenotypes and

EMT in oral cancer cell lines (247). Nicotine is believed to produce these effects through its action on nicotinic acetylcholine receptors (nAChRs). Furthermore, in oral cancer cells, the pro-invasive effects of nicotine are mediated via α7-nAChR. Thus, expression of these nAChRs was evaluated for smoke and radiation-treated cells **(Appendix K)**. Overall, most nAChRs were down-regulated in cell treated with cigarette smoke, suggesting that over-activation of these receptors caused their down-regulation. However, a small number of nAChRs, including α5, α6, α7 and β1 were up-regulated in the smoke treated cells. Furthermore, both radiation and the combination treatment caused even higher up-regulation of α7-nAChR. Thus, cigarette smoke is likely to induce a more invasive phenotype in OPC cells through the action of α7-nAChR.

Radiation Treatment

Radiation treatment also promoted the expression of genes involved in migration, cell adhesion and angiogenesis. The top 9 up-regulated genes were involved in response to inflammatory cytokines such as TNFα, followed by strong up-regulation of p21 cell-cycle inhibitor (CDKN1A). The most prominently up-regulated pro-invasive genes included ITGA5, ADAM8, MMP1, and CD44. This is in support of other studies which have found that radiation can promote de-differentiation of non-stem head and neck cancer cells into cancer stem cells (172).

ITGA5 is an oncogenic integrin with down-stream signalling that promotes cell proliferation and migration. Indeed, the findings of ITGA5 overexpression in these cells are in support of a previous study which found that ITGA5 overexpression promoted migration and invasion in oral cancer cells, as well as promoting the expression of EMT markers SNAI1 and VIM (248).

Combination Treatment of Cigarette Smoke and Radiation

The real interest of this investigation was in the effects of combination treatment of cigarette smoke and radiation, and whether any synergism existed in the promotion of invasive and stem-like phenotypes. As expected, this combination treatment caused the highest number of differentially expressed genes. **Figure 5.3** shows that many of these genes were shared with the smoke and radiation treatments alone. As such, pathway enrichment found many of the same pathways to be affected, namely, inflammatory cytokine signalling, cell adhesion, migration, and angiogenesis **(Figure 5.4)**. When differential expression was analysed between combination treatment and radiation alone, these same pathways were present again, indicating that synergism of this pathway up-regulation was occurring **(Figure 5.5)**.

A number of specific genes that were affected by smoke or radiation alone also showed enhanced effects in the combination treatment. Some of these can be visualised clearly in the heatmaps **(Figure 5.6)**, while others were retrieved from the differential expression data (**Appendix J**).

Genes that showed the highest synergy included many invasion-related genes. For example, PTGS2, or COX-2, is an enzyme involved in prostaglandin biosynthesis which has been shown to promote motility

and invasion in breast cancer cells (249). Its expression was increased 7-fold and 56-fold following smoke and radiation treatment respectively, but 79-fold after combination treatment. Additionally, MMP9 expression was 4-fold and 27-fold higher after smoke or radiation, but 62-fold higher after combination treatment. As another example, ITGA5, the pro-proliferative and pro-invasive integrin discussed previously was increased 2.3-fold and 8-fold after smoke or radiation, but 11-fold after combination treatment. Finally, the key pro-angiogenesis factor VEGFA also showed synergism, with 1.6 and 3.2-fold increases after smoke or radiation, but a 4-fold increase with combination treatment.

What was interesting, however, was a large number of genes that were uniquely differentially expressed only after combination treatment, and did not appear to be affected by each treatment individually. Pathway analysis showed that these genes were mostly involved in metabolic pathways, which could suggest that these cells are more metabolically active following radiation treatment, however this hypothesis is contradicted by the down-regulation of cell cycle pathways. Some uniquely up-regulated genes were relevant to invasion, such as the secreted protease cathepsin A which was increased 1.8-fold after combination treatment, claudin 4 (2.5-fold increase), and Secretory leukocyte protease inhibitor (SLPI) (4-fold increase). SLPI has recently been identified as a metastasis-promoting gene which acts through the FoxM1 pathway (250). Thus, it is unclear whether cells that receive combination treatment are more resistant and likely to resist radiation, however it is clear that they uniquely express pro-invasive genes.

Finally, some important genes were also synergistically repressed by combination treatment of smoke and radiation. For example, the two most strongly down-regulated genes in the combination group were keratin 5 and tensin 3 (TNS3). Keratin 5 is used as a marker of EMT, where it is a marker of epithelial cells rather than mesenchymal cells, and thus its down-regulation supports EMT (251). Tensin 3 has been previously identified as a negative regulator of cell migration in kidney cancer (252).

5.5 Conclusions

Radiation treatment and smoking are known to promote cancer metastasis. Previously, this was suggested to occur via *in vivo*-specific effects of hypoxia. It has been shown here that both radiation and smoke are able to promote invasiveness, EMT, and angiogenesis *in vitro*, with both observable phenotypic changes and supportive changes in gene expression. Additionally, in many cases, radiation and smoke act synergistically to alter markers of cancer invasiveness. This may be mediated through nicotine activity on the α7nAChR, which is also synergistically up-regulated in response to radiation and smoke treatment.

Chapter 6: General Discussion

6.1 Introduction

Cancers are a major cause of illness, both in Australia, and globally. However, due to improved understanding, the capacity for earlier detection, and improved therapeutics, the overall cancer incidence and death rates have both been in decline, showing a decrease of 24% since 1982. Nevertheless, a select few cancer types have demonstrated an opposite trend over this time period in Australia, and their incidence and associated mortality has been on the rise. Liver cancer in particular, and cancers of the oesophagus have had alarming mortality rate increases of 204% and 6.8% respectively (2).

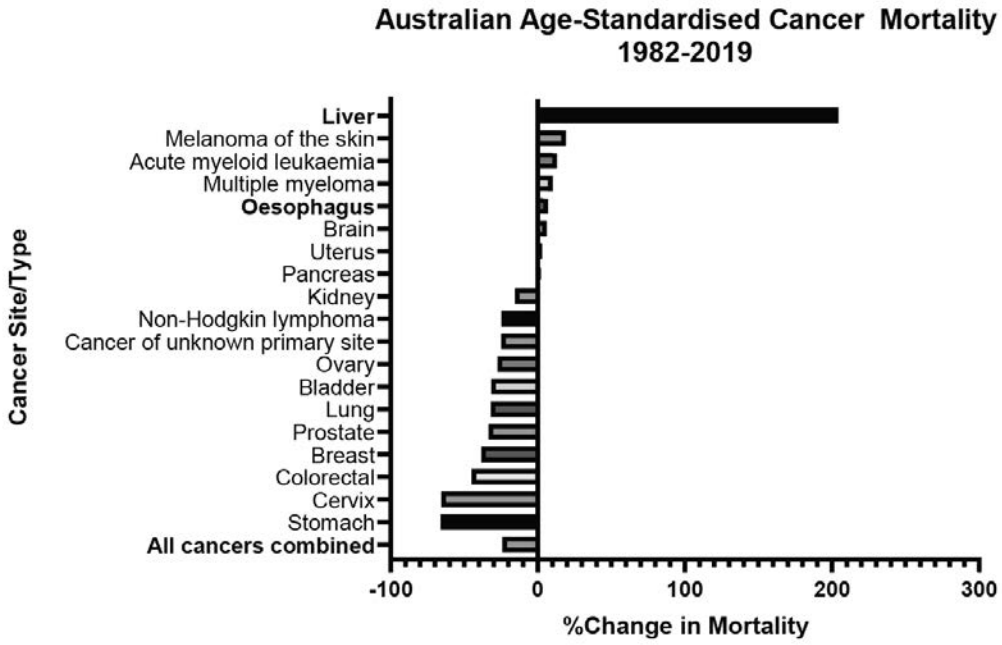

Figure 6.1: Estimated percentage change in age-standardised mortality rates for selected cancers between 1982 and 2019. Data was retrieved from the Australian Institute of Health and Welfare 2019 Cancer Report (2).

For this reason, innovative strategies are required to treat these cancer types. Thus, this thesis aimed to investigate and characterise a novel gene called HORMAD2 in liver cancer, and evaluate whether it could be a viable target for future treatment approaches. Furthermore, a bioinformatic analysis of genetic pathways in OPC was used to evaluate concerns surrounding current radiation-based approaches to treatment of this cancer in Australia.

6.2 Liver Cancer

The primary form of liver cancer is HCC, and it can develop as a result of viral hepatic infections, exposure to toxins such as Aflatoxin, and as a progressive result of non-alcoholic fatty liver disease-induced inflammation. Increasingly, NAFLD-induced HCC is becoming the major contributor to the HCC burden, thanks to improved control of hepatitis B and C viruses and improved sanitation (12). The increasing prevalence of NAFLD can easily be linked with the increasing prevalence of obesity (253), which in turn has been repeatedly linked to the advent of high-fructose corn syrup into the 'western diet'.

Fructose metabolism differs from normal glucose metabolism in a number of key ways. For instance, fructose metabolism is not regulated by circulating insulin levels because it does not promote insulin release. Additionally, fructose metabolism within a cell skips the key regulatory steps of glucose metabolism that prevent the build-up of down-stream molecules like ATP, and thus much more rapidly promote the synthesis of fatty acids. Furthermore, fructose has been linked to the onset of oxidative-stress, which promotes inflammation and DNA-damaging ROS (254). Based on these known pathogenic pathways, the previously discussed preliminary data lead to the discovery of HORMAD2 as a novel target in HCC.

6.2.1 HORMAD2

HORMAD2 is a poorly characterised meiotic protein with only a handful of published articles describing its function, all of which are limited to meiosis (32-34, 36). HORMAD2 is known to accumulate on asynaptic chromosomes during meiosis, which results in the elimination of spermatocytes and oocytes with asynaptic chromosomes (32-34). However, HORMAD2 is also expressed in the liver – expression which is down-regulated in advancing NAFLD (50) and in adiponectin-knockout mice fed a high-fructose diet when compared with those on a normal diet **(Figure 1.1)**.

HORMAD2 has been previously identified as a cancer testis antigen (CTA) in lung cancer (40). CTA genes are normally only expressed in the testes, but are aberrantly expressed in tumour tissue, making them great gene targets for cancer therapies as their function in the target organs are, by definition, dispensable, minimising the chances of side-effects. HORMAD2 was given this delineation based on aberrant expression in lung cancer (40). However, the opposite appears to be true in the liver, where HORMAD2 is expressed in healthy liver tissue.

6.2.2 HORMAD2's Role in HCC is Likely Tumour Suppressive

As discussed previously, HORMAD2 has been shown to be expressed in normal liver tissue, and the progression of NAFLD involves decreased HORMAD2 expression (50). The results of the *in silico* analysis of HORMAD2 in human HCC herein support this. HORMAD2 expression is higher in more

obese patients **(Figure 3.3)**, but its expression is consistently lost in HCC tumour tissue and this decrease becomes stronger at more advanced stages of HCC **(Figure 3.2 and Figure 3.3)**. The higher expression in more obese patients suggests that HORMAD2 is up-regulated to combat some cancer-promoting characteristic of obesity, such as chronic inflammation, or metabolic or oxidative stress. Furthermore, the loss of HORMAD2 expression in HCC, and progressive decrease throughout cancer severity suggests that HORMAD2 down-regulation is important for HCC progression.

6.2.3 HORMAD2 is Regulated by Several Factors

6.2.3.1 DNA-Methylation

If HORMAD2 is indeed a potential therapeutic target as a tumour suppressor, it is important to understand how its expression is regulated in cancer so that this can be manipulated. HORMAD2 expression was significantly associated with DNA methylation at a collection of CpG sites mostly within the HORMAD2 promoter (immediately upstream) **(Figure 3.11)**. Thus, hypermethylation of the HORMAD2 promoter in HCC likely plays an important role in its repression.

Interestingly, hypermethylated CpG sites were found to fall within a predicted binding site for the p53 tumour suppressor. p53 is an apical DNA-damage response protein and a key regulator of essentially every DNA-repair pathway and it also controls cell-cycle checkpoints in response to DNA damage (154). Additionally, p53 is made inactive by a genetic mutation in more than 50% of HCC cases (153), which could further explain the loss of HORMAD2 expression in HCC if it is indeed regulated by this transcription factor (**Figure 6.2**). This finding also provides support for HORMAD2 being involved in the DNA-damage response.

6.2.3.2 miRNAs

HORMAD2 was also found to be inversely correlated with a number of miRNAs. Indeed, almost all of these micro-RNAs have been shown to influence HCC progression in some way based on previous publications, which makes it difficult to narrow down likely candidates and analyse the data any further. However, one key candidate for HORMAD2-regulation is mir-182, which had the 7th strongest negative correlation with HORMAD2 expression **(Figure 3.12)**. The reason why this miRNA is of particular interest is because it has been shown to be 'sponged' by the long non-coding RNA LOC255167 (224), meaning that LOC255167 sequesters its activity, and LOC255167 was also the most strongly positively-correlated gene in the genome with HORMAD2 expression. mir-182 is a known oncogene which promotes metastasis and cell proliferation in HCC through activation of AKT/FOXO and WNT/β-catenin signalling (224, 225).

6.2.3.3 Metabolic Factors

In vivo **(Figure 1.1)** and *in vitro* **(Figure 4.1)** data both suggest that fructose down-regulates HORMAD2. This effect may be non-causative, because fructose is also toxic to HCC cells, which

means that HORMAD2 may be down-regulated indirectly as part of the cell-cycle shut-down. However, fructose has been shown to regulate the expression of many important genes, including repression of PPARα and up-regulation of pro-inflammatory cytokines such as TNFα and IL-6. While the mechanisms of fructose-regulated gene expression are not understood, fructose appears to affect CpG methylation (20). As well as gene expression, fructose delivery has also been shown to attenuate insulin-signalling by decreasing phosphorylation of the insulin receptor substrates -1 and -2 (IRS-1, IRS-2) (233).

6.2.3.4 Nuclear Localisation

In HCC cell lines, HORMAD2 was found to localise in the nucleus **(Figure 4.5)** as it does in the testes (32, 36). This suggests that some of HORMAD2's known nuclear functions in the testes may be upheld in the liver. It was also noted however, that a significant amount of HORMAD2 protein remained in the cytoplasm. This is to be expected due to the fact that proteins are created in the cytoplasm before being transported into the nucleus. However, it is interesting to note that the HORMAD2 nuclear localisation signal is in its C-terminal tail, in close proximity to both its HORMA-domain closure motif as well as an S/T-Q site for phosphorylation by PIKK-family kinases. Thus, it is interesting to consider that another component of HORMAD2 regulation could be sequestration to the cytoplasm by either interaction with another HORMA domain or phosphorylation of this S/T-Q site, either of which could theoretically impact the binding of nuclear-transport proteins.

6.2.4 Proposed Functions of HORMAD2 in HCC

HORMAD2 is known to be recruited to asynaptic chromosomes in meiosis and mediate the recruitment of ATR to these chromosomes, an important DDR protein. Outside of meiosis, HORMAD2's function remains unelucidated. However, given these known functions, it was hypothesised that HORMAD2 is likely to play a role in the DDR.

The two main features of the DDR pathway are the repair of damaged-DNA, and the stalling of the cell cycle or even induction of apoptosis. Outside of sex tissues, HORMAD2 expression appears to be mainly limited to the liver **(Figure 1.7)** (49). One possible explanation for this is because the liver is prone to constant insult of DNA damage **(see Chapter 1.3.4)** (255), and also because the liver is a regenerative organ capable of extensive wound-healing. These two factors make it prone to the accumulation of genetic errors, and it may be that additional systems are necessary to restrict the survival and proliferation of genetically compromised cells.

The following section will outline the proposed roles of HORMAD2 in the liver. These proposed functions are simply hypotheses based on previous literature, and on the results of the various data presented in this thesis. The hypothesised molecular pathways of these functions are summarised in **Figure 6.2**.

6.2.4.1 Response to oxidative stress

Evidence suggests that HORMAD2 is up-regulated in response to metabolic and oxidative-stress, most likely in the form of oxidative DNA damage (**Figure 6.2**). For example, *in silico* analysis found that HORMAD2 expression is positively correlated with many members of the cytochrome P450 complex **(Figure 3.7)**, which are known endogenous sources of DNA-damaging and oxidative stress-inducing ROS (67). Additionally, data-mining of other gene expression datasets found that HORMAD2 was up-regulated in human hepatocytes after 48 hour treatment with valproic acid (256), which induces liver steatosis and ROS production in hepatocytes (257) (258). Unfortunately, no *in vitro* data was acquired to support this.

6.2.4.2 Cell Cycle Control

It also appears possible that HORMAD2 down-regulates the cell cycle. *In silico* data in support of this showed that HORMAD2 expression was strongly negatively correlated with cell-cycle pathways and nucleic acid synthesis. While laboratory observations supported this *in vitro* while generating stable cell lines, no *in vitro* data acquired supported this hypothesis. This could be explained if HORMAD2 expression is somehow buffered in HCC cells, and although its expression may transiently be altered, the cells are able to recover after passaging. This could be investigated in future through the use of transient over-expression and knockdown systems, however this unfortunately was not achieved in this project. Some *in vitro* data which did support this hypothesis however was the tumoursphere assay, which showed that knockdown of HORMAD2 allowed the formation of much larger and aggregated spheroids. One possible reason for this effect will be discussed later in relation to the addition of insulin into the cell culture media for this experiment.

6.2.4.3 DNA-Repair

It remains unclear how HORMAD2 affects DNA repair. In comet assays, knockdown of HORMAD2 was associated with improved repair of DNA after both H_2O_2 treatment and radiation treatment. As discussed previously, this could be explained by HORMAD2's potential role in inter-homolog bias (25, 32). Briefly, somatic cells and meiocytes differ in the repair of double-stranded breaks by their recombination-partner choice. In normal cells, the ideal candidate for template-directed repair of DSBs is the sister chromatid, because it is theoretically a perfect template. However, in meiocytes repair with the sister chromatid is inhibited to create a bias towards the homologous chromosome and thus facilitate crossover events (39). If this function were to carry over into a somatic cell such as in the liver, it could translate to inhibition of double-stranded break repair via homologous recombination.

In cell viability assays however, knockdown or overexpression of HORMAD2 did not cause a strong, nor consistent effect on cell survival after treatment with DNA-damaging agents. Thus, it appears that HORMAD2 in these experiments attenuates repair of DNA in the short term (1 hour), but does not have a significant effect on the overall capacity of the cells to repair and survive after longer time frames (24

and 48 hours). In contrast to the previous paragraph then, this puts HORMAD2's suppression of repair more in line with NHEJ, which occurs in around 30 minutes, rather than HRR, which takes around 7 hours (259).

Additionally, a high-throughput protein-protein interaction study identified NSD2 as a binding partner of HORMAD2. NSD2 is an oncogenic protein methylase that methylates many proteins including histones, p53 and PTEN in response to DNA damage, and this methylation site mediates recruitment of 53BP1, the key promoter of NHEJ pathway choice (217, 229, 230). This implicates HORMAD2 in promoting NHEJ, which is contradictory to the *in vitro* data.

Finally, turning to the literature for answers also gives contradictory findings. HORMAD1 is a closely related and functionally similar protein to HORMAD2 (32-34) and has also been identified as a cancer-testis antigen. Two independent studies have found that HORMAD1 is over-expressed in lung cancer, however one study concluded that HORMAD1 is an inhibitor of HRR (47), while the other concluded that it promotes HRR (48). Thus, assuming that HORMAD1 and HORMAD2 would function similarly in cancer, this presents further confusion to HORMAD2's role in HCC, and hence, further experimentation is required to resolve these many contradictions.

6.2.4.4 Insulin-Signalling Attenuation

Multiple lines of evidence also support a potential role of HORMAD2 in attenuation of insulin signalling. The insulin signalling pathway promotes lipid and glycogen synthesis, but also promotes cell proliferation via MAPK signalling (260). A close relative of HORMAD2 is MAD2, another HORMA-domain-containing protein which is involved in the mitotic spindle-assembly checkpoint. Thus, it is similar to HORMAD2 in both structure and function. Recently, it has been discovered that MAD2 is capable of mediating endocytosis of the insulin receptor and thus down-regulating its activity (261). Given the fact that HORMA domain proteins are known to interact with each other and function together, it is possible that HORMAD2 could also cooperate in this pathway.

This possibility is supported by the findings of this thesis. In the *in silico* analysis of HCC patients, the genes with strongest positive correlation with HORMAD2 were involved in lipid oxidation and gluconeogenesis **(Figure 3.7)** (appendix), all processes that are down-regulated by insulin signalling. Thus, high expression of HORMAD2 could mediate attenuation of the insulin-mediated repression of these genes. Additionally, *in vitro* knockdown of HORMAD2 only showed a strong effect on HCC cell growth in the tumoursphere assay, which differed from other growth assays performed by the inclusion of insulin in the cell culture media. In this assay, HORMAD2-knockdown led to the development of much larger and more aggregative spheroids in 3-dimensional culture. This suggests that the prominent HORMAD2 expression in the Huh7 cell line may have been attenuating the insulin-mediated activation of MAPK, whereas knockdown of HORMAD2 removed this attenuation (**Figure 6.2**).

6.2.4.5 Regulation of Stem Cell Number

CSCs are cancer cells that have acquired stem-like properties which give them a unique capacity for self-renewal. CSCs therefore play an important role in cancer metastasis, and are also more radio- and chemo-resistant (185, 186). Tumoursphere assays, like those performed in this investigation **(Figure 4.9)** can be used to detect CSCs, because non-CSCs are incapable of proliferation in non-adherent conditions (237). Because HORMAD2 knockdown promoted aggregated spheroid formation in the tumour sphere assay, it is possible that HORMAD2 could be a negative regulator of CSCs. While the mechanism behind this remains to be elucidated, it could be related to HORMAD2's functional interactions with BRCA1, which is also a known negative regulator of CSCs (238, 239). The Hebbard lab is actively investigating this possibility further.

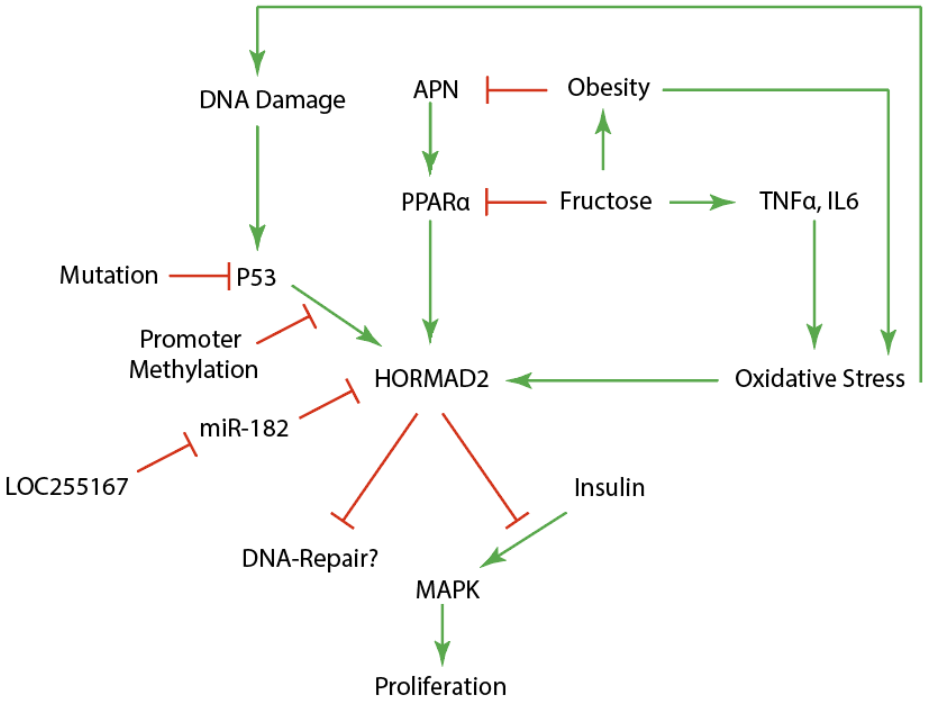

Figure 6.2: Proposed pathway of HORMAD2 signalling in the liver. HORMAD2's function in the liver remains unknown, and the nature of its regulation requires experimental confirmation. However, HORMAD2 activity may impair cell proliferation through attenuation of insulin signalling similarly to MAD2, and may also attenuate DNA repair in response to oxidative stress in the liver. Bioinformatic analysis suggests that HORMAD2 may be regulated by promoter methylation in the binding sites of P53, as we as regulation by miR-182. Green arrows indicate up-regulation or activation, while red arrows indicate down-regulation or inhibition. APN, adiponectin; TNFα, tumour-necrosis factor alpha; PPARα, Peroxisome proliferator-activated receptor alpha; IL6, interleukin 6; MAPK, mitogen-activated protein kinase

6.2.5 Limitations

The results of this project were limited by a number of factors. Firstly, the TCGA-LIHC dataset analysed for *in silico* characterisation of HORMAD2 was composed of primarily viral-associated HCC, which made it a non-ideal dataset for the study of HORMAD2, given that it was hypothesised to play a

role in NAFLD-associated HCC. On top of this, most of the bioinformatic analyses performed for both the TCGA-LIHC cohort and the oral cancer cell culture experiments are largely correlative. Thus, the results of these analysis are not reinforced by any evidence of causation and therefore require further experimental investigation to support their interpretation.

Secondly, the *in vitro* studies were hindered in several ways which need to be considered when evaluating the significance of their findings. The experiments were performed on modified cell lines which demonstrated very weak overexpression and knockdown of HORMAD2, which is likely to have influenced the lack of phenotypic effects. Additionally, reproducibility of this data was greatly impaired by a lack of access to high quality reagents, particularly antibodies, due to cost and time restraints and difficulties associated with the COVID-19 pandemic. As a result, many of the findings in chapter 4 are unfortunately based on inadequate sample sizes to form strong conclusions and, given access to further resources, it is imperative that these experiments be reproduced.

Thirdly, given the finding of potential interactions with insulin signalling, it is possible that the phenotypic effects of HORMAD2 require an *in vivo* model with in-tact cell-cell signalling systems to observe their effect. This could explain why many of the findings in the human bioinformatic dataset did not translate into the *in vitro* data. Thus, future *in vivo* investigations in a suitable model may be required to truly elucidate the functions of HORMAD2.

6.2.6 Conclusion

This investigation has greatly contributed to the understanding of HORMAD2 in HCC. However, many questions remain unanswered, and much more experimentation is required to elucidate the complexity of HORMAD2's involvement. It is here proposed that HORMAD2 in the liver does not act as a classic cancer testis antigen, but it instead plays a real functional role in the normal healthy liver. HORMAD2 is a HCC tumour suppressor which may down-regulate both DNA repair and the cell cycle in response to oxidative and metabolic stress. This is likely to prevent the continual repair and survival of damaged cells. Cell cycle control could be achieved through attenuation of insulin-mediated MAPK activation. Furthermore, down-regulation of HORMAD2 by fructose, aberrant expression of oncogenic micro RNAs, and promoter hypermethylation may contribute to the pathogenesis of obesity and NAFLD-induced HCC.

6.3 OPC

The primary risk factors for OPC are smoking and HPV infection. HPV negative oral cancers are known to be associated with worse survival and increased occurrence of metastasis (170-172). Currently, OPC is treated usually either through resection of the tumour, radiation or both, and over the past few decades, little progress has been made in improving overall survival.

It is well known that radiation can be both beneficial and detrimental to cancer metastasis. On one hand, it has been known for many decades that sub-curative radiation therapy promotes metastasis (176, 177). Additionally, during radiation therapy cancers enter a phase of accelerated repopulation, increasing their proliferative capacity. Consequently, it is crucial that a course of curative intent radiation therapy is completed, or not embarked upon in the first place. It has been suggested that this radiation-associated increased metastasis is due to disruption of the tumour-associated blood vessels, which allows an increased influx of tumour cells into the bloodstream (175). However, it has also been shown that radiation treatment of a primary tumour can cause shrinkage of secondary tumours at distant sites, through a phenomenon known as the abscopal effect. While the mechanisms of this are poorly understood, it is believed that this connection is made through immune mechanisms, however this is outside the scope of this study (175).

Similarly, continued use of tobacco smoke during radiation therapy has also been shown to increase the risk of metastasis. As with radiation, it has been suggested that the most likely explanation for this is the effect of smoking on chronic hypoxia (171). Low levels of oxygen, known as hypoxia, is unavoidable in solid tumours, and is known to lead to more aggressive, radioresistant cancer phenotypes and also promotes angiogenesis (175). Thus, it is believed that because smoking promotes hypoxia, which in turn promotes angiogenesis and more radioresistant phenotypes, there is in an indirect increased metastatic potential of these tumour cells.

However, there are multiple lines of evidence to suggest that both radiation and cigarette smoke can directly promote more invasive phenotypes *in vitro*. Thus, the aim of this study was to investigate *in vitro* the effect of cigarette smoke and radiation on OPC cells, and to elucidate the molecular pathways involved in these phenotypic changes.

6.3.1 Radiation and smoke act synergistically to promote invasion

FaDu OPC cells were treated with cell culture media infused with cigarette smoke, radiation, or both. Radiated cells received 1.8 Gy fractions of radiation totalling 27 Gy over 3 weeks. This dosage is representative of the treatment prescribed to patients with microscopic disease. RNA was then isolated from these cells and next-generation sequencing was performed.

All treatments promoted the expression of genetic pathways involved in locomotion, cellular adhesion, and angiogenesis. All treatments also caused an expected amount of cell death, which was seen in down-regulation of cell-cycle pathways in the sequencing analysis. Immediately, this suggests that both radiation and smoke treatment directly promote the expression of invasion-related genes *in vitro*.

Of particular interest were the genes on which smoke and radiation had a synergistic effect. COX-2, multiple MMPs, ITGA5, and VEGFA, B, and C, and CD44 were all increased by both smoke and radiation treatment, and showed synergistic increase after combination treatment, while invasion-

repressive genes such as KRT5 and TNS3 were synergistically down-regulated. Additionally, some genes were almost completely unaltered by radiation or smoke treatment alone, but were strongly up-regulated after combination treatment. Genes with this pattern included CTSA and CTSD, CLDN4, and SLPI. Thus, this encompasses up-regulation of multiple secreted proteases which are important for invasion, markers of EMT and cancer stem cells, and angiogenesis-promoting growth factors. Some less well-curated of these genes include TNS3, which has been previously identified as a negative regulator of cell migration in kidney cancer (252), and the pro-invasive SLPI gene (250) **(Appendix J)**.

Not all changes in gene expression were consistent with the *in vitro* observations however, and cannot be completely explained. For example, the core regulators of cell motility RhoA and Rac1 were both down-regulated after either treatment. Looking more in depth at the gene expression however, it is apparent that other GTPases such as RhoB, and RhoC were in fact increased 2.7-fold and 1.7-fold in cells that received combination treatment. Although not listed in the curated genes set, RhoC is also associated with metastasis in multiple cancer types and seems to more consistently be involved in metastasis than RhoA. Indeed, though not as strong a trend, RhoC was synergistically up-regulated by smoke and radiation. The literature is in debate as to whether RhoB is an oncogene or tumour suppressor, however it has been suggested that it could be important in some cancer types (262) **(Appendix J)**. Additionally, the many ALDH enzymes were very variable, with some being up-regulated and some being down-regulated in response to radiation and smoke. ALDH enzymes are important regulators of cell stemness, and while previous studies have shown that specific ALDH isoforms may be up or down-regulated in cancer, an explanation for the specific pattern of expression in this study could not be identified (263).

The mechanism remains to be elucidated as to how cigarette smoke can induce the genetic changes observed in this experiment. However, one promising candidate is the α7 nicotinic acetylcholine receptor. This receptor has been previously shown to be important in the regulation of nicotine-induced invasiveness (247). Additionally, in this study it was found that while many nAChRs are down-regulated in response to cigarette smoke, α7nAChR appears to be synergistically increased by both radiation and cigarette smoke as discussed for many of these effector genes.

6.3.2 Implications

Overall, the results of high-throughput sequencing analysis in combination with the clinical and *in vitro* observations strongly suggest that both smoke and radiation can induce more invasive phenotypes of oral cancer. Radiation treatment is known to promote tumour invasiveness, which is why it is important that radiation treatment is followed through and that the tumour is destroyed before invasion occurs. The importance of this investigation is that it shows that patients undergoing radiation therapy who continue to smoke may be at an increased risk of metastasis during their radiation treatment. Although it may seem simple to suggest that patients should be encouraged not to smoke during radiation

treatment, research suggests that such patients are highly likely to continue to smoke (264). Thus, genes, such as the ones identified in this study which are synergistically up-regulated by these invasion-promoting factors of radiation and smoke, are extremely valuable targets for the development of future conjoint therapies for patients undergoing radiation treatment.

6.3.3 Limitations

There are several factors that limited the efficacy of this study. For example, it is known that micro RNAs play an important role in EMT and radioresistance of head and neck cancers. Such micro RNAs were not studied in this investigation, and therefore represent an important knowledge gap for future studies. Additionally, this investigation was only performed on a limited scale, with a small sample size (n = 3 per treatment) and only in a single cell line. Cancers are extremely heterogeneous by nature, and thus a single cell type can never replicate the complexity of an *in vivo* solid tumour, but in can allow us to begin to tease apart the effects of a specific treatment.

6.3.4 Conclusion

Nonetheless, this study represents an invaluable resource for the future development of conjoint therapies for patients who continue to smoke while undergoing radiation therapy for OPC. It is likely that these patients are at an increased risk of metastasis, but by developing inhibitors for the known cellular pathways involved in this increased metastasis, this risk may be able to be mitigated in future.

References

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144(5):646-74.
2. Health AIo, Welfare. Cancer in Australia 2019. Canberra: AIHW; 2019.
3. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer. 2015;136(5):E359-E86.
4. Hashim D, Boffetta P, La Vecchia C, Rota M, Bertuccio P, Malvezzi M, et al. The global decrease in cancer mortality: trends and disparities. Annals of Oncology. 2016;27(5):926-33.
5. Altekruse SF, Devesa SS, Dickie LA, McGlynn KA, Kleiner DE. Histological classification of liver and intrahepatic bile duct cancers in SEER registries. Journal of registry management. 2011;38(4):201-5.
6. Massarweh NN, El-Serag HB. Epidemiology of Hepatocellular Carcinoma and Intrahepatic Cholangiocarcinoma. Cancer control : journal of the Moffitt Cancer Center. 2017;24(3):1073274817729245-.
7. Bataller R, Brenner DA. Liver fibrosis. J Clin Invest. 2005;115(2):209-18.
8. Rawla P, Sunkara T, Muralidharan P, Raj JP. Update in global trends and aetiology of hepatocellular carcinoma. Wspolczesna Onkologia. 2018;22(3):141-50.
9. Global Burden of Disease Liver Cancer C, Akinyemiju T, Abera S, Ahmed M, Alam N, Alemayohu MA, et al. The Burden of Primary Liver Cancer and Underlying Etiologies From 1990 to 2015 at the Global, Regional, and National Level: Results From the Global Burden of Disease Study 2015. JAMA oncology. 2017;3(12):1683-91.
10. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. Journal of Hepatology. 2019;70(1):151-71.
11. Cheng A-L, Kang Y-K, Chen Z, Tsao C-J, Qin S, Kim JS, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. The Lancet Oncology. 2009;10(1):25-34.
12. Goldberg D, Ditah IC, Saeian K, Lalehzari M, Aronsohn A, Gorospe EC, et al. Changes in the Prevalence of Hepatitis C Virus Infection, Nonalcoholic Steatohepatitis, and Alcoholic Liver Disease Among Patients With Cirrhosis or Liver Failure on the Waitlist for Liver Transplantation. Gastroenterology. 2017;152(5):1090-9.e1.
13. Piñero F, Pages J, Marciano S, Fernández N, Silva J, Anders M, et al. Fatty liver disease, an emerging etiology of hepatocellular carcinoma in Argentina. World journal of hepatology. 2018;10(1):41-50.
14. Piscaglia F, Svegliati-Baroni G, Barchetti A, Pecorelli A, Marinelli S, Tiribelli C, et al. Clinical patterns of hepatocellular carcinoma in nonalcoholic fatty liver disease: A multicenter prospective study. Hepatology (Baltimore, Md). 2016;63(3):827-38.
15. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, Obesity, and Mortality from Cancer in a Prospectively Studied Cohort of U.S. Adults. New England Journal of Medicine. 2003;348(17):1625-38.
16. Polyzos SA, Kountouras J, Mantzoros CS. Adipokines in nonalcoholic fatty liver disease. Metabolism. 2016;65(8):1062-79.
17. Park EJ, Lee JH, Yu GY, He G, Ali SR, Holzer RG, et al. Dietary and genetic obesity promote liver inflammation and tumorigenesis by enhancing IL-6 and TNF expression. Cell. 2010;140(2):197-208.
18. Dewdney B, Roberts A, Qiao L, George J, Hebbard L. A Sweet Connection? Fructose's Role in Hepatocellular Carcinoma. Biomolecules. 2020;10(4).
19. DiNicolantonio JJ, Mehta V, Onkaramurthy N, O'Keefe JH. Fructose-induced inflammation and increased cortisol: A new mechanism for how sugar induces visceral adiposity. Progress in cardiovascular diseases. 2018;61(1):3-9.
20. DiStefano JK. Fructose-mediated effects on gene expression and epigenetic mechanisms associated with NAFLD pathogenesis. Cell Mol Life Sci. 2020;77(11):2079-90.

21. Softic S, Gupta MK, Wang GX, Fujisaka S, O'Neill BT, Rao TN, et al. Divergent effects of glucose and fructose on hepatic lipogenesis and insulin signaling. The Journal of clinical investigation. 2017;127(11):4059-74.
22. Elliott SS, Keim NL, Stern JS, Teff K, Havel PJ. Fructose, weight gain, and the insulin resistance syndrome. The American Journal of Clinical Nutrition. 2002;76(5):911-22.
23. Walker S, Wankell M, Ho V, White R, Deo N, Devine C, et al. Targeting mTOR and Src restricts hepatocellular carcinoma growth in a novel murine liver cancer model. PloS one. 2019;14(2):e0212860.
24. Aravind L, Koonin EV. The HORMA domain: a common structural denominator in mitotic checkpoints, chromosome synapsis and DNA repair. Trends in Biochemical Sciences. 1998;23(8):284-6.
25. Kim Y, Rosenberg SC, Kugel CL, Kostow N, Rog O, Davydov V, et al. The chromosome axis controls meiotic events through a hierarchical assembly of HORMA domain proteins. Developmental cell. 2014;31(4):487-502.
26. West AMV, Komives EA, Corbett KD. Conformational dynamics of the Hop1 HORMA domain reveal a common mechanism with the spindle checkpoint protein Mad2. Nucleic acids research. 2018;46(1):279-92.
27. Bates M, Furlong F, Gallagher MF, Spillane CD, McCann A, O'Toole S, et al. Too MAD or not MAD enough: The duplicitous role of the spindle assembly checkpoint protein MAD2 in cancer. Cancer Letters. 2020;469:11-21.
28. Handel MA, Schimenti JC. Genetics of mammalian meiosis: regulation, dynamics and impact on fertility. Nat Rev Genet. 2010;11:124.
29. Turner JMA, Mahadevaiah SK, Fernandez-Capetillo O, Nussenzweig A, Xu X, Deng C-X, et al. Silencing of unsynapsed meiotic chromosomes in the mouse. Nature genetics. 2004;37:41.
30. Fernandez-Capetillo O, Mahadevaiah SK, Celeste A, Romanienko PJ, Camerini-Otero RD, Bonner WM, et al. H2AX Is Required for Chromatin Remodeling and Inactivation of Sex Chromosomes in Male Mouse Meiosis. Developmental cell. 2003;4(4):497-508.
31. Rinaldi VD, Bolcun-Filas E, Kogo H, Kurahashi H, Schimenti JC. The DNA Damage Checkpoint Eliminates Mouse Oocytes with Chromosome Synapsis Failure. Molecular cell. 2017;67(6):1026-36 e2.
32. Wojtasz L, Daniel K, Roig I, Bolcun-Filas E, Xu H, Boonsanay V, et al. Mouse HORMAD1 and HORMAD2, two conserved meiotic chromosomal proteins, are depleted from synapsed chromosome axes with the help of TRIP13 AAA-ATPase. PLoS genetics. 2009;5(10):e1000702.
33. Kogo H, Tsutsumi M, Inagaki H, Ohye T, Kiyonari H, Kurahashi H. HORMAD2 is essential for synapsis surveillance during meiotic prophase via the recruitment of ATR activity. Genes to cells : devoted to molecular & cellular mechanisms. 2012;17(11):897-912.
34. Wojtasz L, Cloutier JM, Baumann M, Daniel K, Varga J, Fu J, et al. Meiotic DNA double-strand breaks and chromosome asynapsis in mice are monitored by distinct HORMAD2-independent and -dependent mechanisms. Genes & development. 2012;26(9):958-73.
35. Royo H, Prosser H, Ruzankina Y, Mahadevaiah SK, Cloutier JM, Baumann M, et al. ATR acts stage specifically to regulate multiple aspects of mammalian meiotic silencing. Genes & development. 2013;27(13):1484-94.
36. Fukuda T, Pratto F, Schimenti JC, Turner JM, Camerini-Otero RD, Hoog C. Phosphorylation of chromosome core components may serve as axis marks for the status of chromosomal events during mammalian meiosis. PLoS genetics. 2012;8(2):e1002485.
37. Lao JP, Hunter N. Trying to Avoid Your Sister. PLOS Biology. 2010;8(10):e1000519.
38. Hong S, Sung Y, Yu M, Lee M, Kleckner N, Kim KP. The logic and mechanism of homologous recombination partner choice. Molecular cell. 2013;51(4):440-53.
39. Humphryes N, Hochwagen A. A non-sister act: recombination template choice during meiosis. Experimental cell research. 2014;329(1):53-60.
40. Liu M, Chen J, Hu L, Shi X, Zhou Z, Hu Z, et al. HORMAD2/CT46.2, a novel cancer/testis gene, is ectopically expressed in lung cancer tissues. Molecular human reproduction. 2012;18(12):599-604.
41. Rosenberg SC, Corbett KD. The multifaceted roles of the HORMA domain in cellular signaling. The Journal of Cell Biology. 2015;211(4):745.

42. Nielsen AY, Gjerstorff MF. Ectopic Expression of Testis Germ Cell Proteins in Cancer and Its Potential Role in Genomic Instability. International journal of molecular sciences. 2016;17(6):890.
43. Shammas MA, Shmookler Reis RJ, Koley H, Batchu RB, Li C, Munshi NC. Dysfunctional homologous recombination mediates genomic instability and progression in myeloma. Blood. 2009;113(10):2290.
44. Widger A, Mahadevaiah SK, Lange J, ElInati E, Zohren J, Hirota T, et al. ATR is a multifunctional regulator of male mouse meiosis. Nature communications. 2018;9(1):2621-.
45. Zhang K, Tang S, Cao S, Hu L, Pan Y, Ma H, et al. Association of polymorphisms at HORMAD2 and prognosis in advanced non-small-cell lung cancer patients. Cancer epidemiology. 2014;38(4):414-8.
46. Wang X, Tan Y, Cao X, Kim JA, Chen T, Hu Y, et al. Epigenetic activation of HORMAD1 in basal-like breast cancer: role in Rucaparib sensitivity. Oncotarget. 2018;9(53):30115-27.
47. Watkins J, Weekes D, Shah V, Gazinska P, Joshi S, Sidhu B, et al. Genomic Complexity Profiling Reveals That HORMAD1 Overexpression Contributes to Homologous Recombination Deficiency in Triple-Negative Breast Cancers. Cancer Discov. 2015;5(5):488-505.
48. Gao Y, Kardos J, Yang Y, Tamir TY, Mutter-Rottmayer E, Weissman B, et al. The Cancer/Testes (CT) Antigen HORMAD1 promotes Homologous Recombinational DNA Repair and Radioresistance in Lung adenocarcinoma cells. Scientific reports. 2018;8(1):15304.
49. Fagerberg L, Hallstrom BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. Molecular & cellular proteomics : MCP. 2014;13(2):397-406.
50. Ryaboshapkina M, Hammar M. Human hepatic gene expression signature of non-alcoholic fatty liver disease progression, a meta-analysis. Scientific reports. 2017;7(1):12361.
51. Mokdad AA, Lopez AD, Shahraz S, Lozano R, Mokdad AH, Stanaway J, et al. Liver cirrhosis mortality in 187 countries between 1980 and 2010: a systematic analysis. BMC Medicine. 2014;12(1):145.
52. Asrani SK, Larson JJ, Yawn B, Therneau TM, Kim WR. Underestimation of Liver-Related Mortality in the United States. Gastroenterology. 2013;145(2):375-82.e2.
53. Boege Y, Malehmir M, Healy ME, Bettermann K, Lorentzen A, Vucur M, et al. A Dual Role of Caspase-8 in Triggering and Sensing Proliferation-Associated DNA Damage, a Key Determinant of Liver Cancer Development. Cancer cell. 2017;32(3):342-59 e10.
54. Wu X, Gu J, Patt Y, Hassan M, Spitz MR, Beasley RP, et al. Mutagen sensitivity as a susceptibility marker for human hepatocellular carcinoma. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 1998;7(7):567-70.
55. Nickoloff JA, Jones D, Lee SH, Williamson EA, Hromas R. Drugging the Cancers Addicted to DNA Repair. Journal of the National Cancer Institute. 2017;109(11).
56. O'Connor Mark J. Targeting the DNA Damage Response in Cancer. Molecular cell. 2015;60(4):547-60.
57. Liu Y, Beard WA, Shock DD, Prasad R, Hou EW, Wilson SH. DNA Polymerase β and Flap Endonuclease 1 Enzymatic Specificities Sustain DNA Synthesis for Long Patch Base Excision Repair. Journal of Biological Chemistry. 2005;280(5):3665-74.
58. Liu Y, Prasad R, Beard WA, Kedar PS, Hou EW, Shock DD, et al. Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta. The Journal of Biological Chemistry. 2007;282(18):13532-41.
59. Racanelli V, Rehermann B. The liver as an immunological organ. Hepatology (Baltimore, Md). 2006;43(S1):S54-S62.
60. Gao B, Jeong W-I, Tian Z. Liver: An organ with predominant innate immunity. Hepatology (Baltimore, Md). 2008;47(2):729-36.
61. Yu L-X, Schwabe RF. The gut microbiome and liver cancer: mechanisms and clinical translation. Nature Reviews Gastroenterology &Amp; Hepatology. 2017;14:527.
62. Hagen TM, Huang S, Curnutte J, Fowler P, Martinez V, Wehr CM, et al. Extensive oxidative DNA damage in hepatocytes of transgenic mice with chronic active hepatitis destined to develop hepatocellular carcinoma. Proceedings of the National Academy of Sciences of the United States of America. 1994;91(26):12808-12.

63. Benten D, Schulze zur Wiesch J, Sydow K, Koops A, Buggisch P, Böger RH, et al. The transhepatic endotoxin gradient is present despite liver cirrhosis and is attenuated after transjugular portosystemic shunt (TIPS). BMC gastroenterology. 2011;11:107-.
64. Jaiswal M, LaRusso NF, Nishioka N, Nakabeppu Y, Gores GJ. Human Ogg1, a Protein Involved in the Repair of 8-Oxoguanine, Is Inhibited by Nitric Oxide. Cancer Research. 2001;61(17):6388.
65. Cao D, Song X, Che L, Li X, Pilo MG, Vidili G, et al. Both de novo synthetized and exogenous fatty acids support the growth of hepatocellular carcinoma cells. Liver international : official journal of the International Association for the Study of the Liver. 2017;37(1):80-9.
66. Enriquez-Cortina C, Bello-Monroy O, Rosales-Cruz P, Souza V, Miranda RU, Toledo-Perez R, et al. Cholesterol overload in the liver aggravates oxidative stress-mediated DNA damage and accelerates hepatocarcinogenesis. Oncotarget. 2017;8(61):104136-48.
67. Hrycay EG, Bandiera SM. Chapter Two - Involvement of Cytochrome P450 in Reactive Oxygen Species Formation and Cancer. In: Hardwick JP, editor. Advances in Pharmacology. 74: Academic Press; 2015. p. 35-84.
68. Fukui H, Brauner B, Bode JC, Bode C. Plasma endotoxin concentrations in patients with alcoholic and non-alcoholic liver disease: reevaluation with an improved chromogenic assay. J Hepatol. 1991;12(2):162-9.
69. Lin RS, Lee FY, Lee SD, Tsai YT, Lin HC, Lu RH, et al. Endotoxemia in patients with chronic liver diseases: relationship to severity of liver diseases, presence of esophageal varices, and hyperdynamic circulation. J Hepatol. 1995;22(2):165-72.
70. Mutlu EA, Gillevet PM, Rangwala H, Sikaroodi M, Naqvi A, Engen PA, et al. Colonic microbiome is altered in alcoholism. American journal of physiology Gastrointestinal and liver physiology. 2012;302(9):G966-78.
71. Machida K, Cheng KTN, Sung VMH, Shimodaira S, Lindsay KL, Levine AM, et al. Hepatitis C virus induces a mutator phenotype: Enhanced mutations of immunoglobulin and protooncogenes. Proceedings of the National Academy of Sciences. 2004;101(12):4262.
72. Machida K, Cheng KTH, Lai C-K, Jeng K-S, Sung VMH, Lai MMC. Hepatitis C Virus Triggers Mitochondrial Permeability Transition with Production of Reactive Oxygen Species, Leading to DNA Damage and STAT3 Activation. Journal of virology. 2006;80(14):7199.
73. Chen CL, Yang HI, Yang WS, Liu CJ, Chen PJ, You SL, et al. Metabolic factors and risk of hepatocellular carcinoma by chronic hepatitis B/C infection: a follow-up study in Taiwan. Gastroenterology. 2008;135(1):111-21.
74. Selzner N, Selzner M, Odermatt B, Tian Y, Van Rooijen N, Clavien PA. ICAM-1 triggers liver regeneration through leukocyte recruitment and Kupffer cell-dependent release of TNF-alpha/IL-6 in mice. Gastroenterology. 2003;124(3):692-700.
75. Fortini P, Pascucci B, Parlanti E, D'Errico M, Simonelli V, Dogliotti E. 8-Oxoguanine DNA damage: at the crossroad of alternative repair pathways. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2003;531(1):127-39.
76. Kakehashi A, Ishii N, Okuno T, Fujioka M, Gi M, Fukushima S, et al. Progression of Hepatic Adenoma to Carcinoma in Ogg1 Mutant Mice Induced by Phenobarbital. Oxidative Medicine and Cellular Longevity. 2017;2017:16.
77. Vartanian V, Lowell B, Minko IG, Wood TG, Ceci JD, George S, et al. The metabolic syndrome resulting from a knockout of the NEIL1 DNA glycosylase. Proceedings of the National Academy of Sciences of the United States of America. 2006;103(6):1864.
78. Di Maso V, Mediavilla MG, Vascotto C, Lupo F, Baccarani U, Avellini C, et al. Transcriptional Up-Regulation of APE1/Ref-1 in Hepatic Tumor: Role in Hepatocytes Resistance to Oxidative Stress and Apoptosis. PloS one. 2015;10(12):e0143289.
79. Lin YT, Liu W, He Y, Wu YL, Chen WN, Lin XJ, et al. Hepatitis B Virus X Protein Increases 8-Oxo-7,8-Dihydro-2´-Deoxyguanosine (8-Oxodg) Level via Repressing MTH1/ MTH2 Expression in Hepatocytes. Cellular Physiology and Biochemistry. 2018;51(1):80-96.
80. Yu Z, Wang R, Chen F, Wang J, Huang X. Five Novel Oncogenic Signatures Could Be Utilized as AFP-Related Diagnostic Biomarkers for Hepatocellular Carcinoma Based on Next-Generation Sequencing. Digestive Diseases and Sciences. 2018;63(4):945-57.

81. Yuan T, Wei J, Luo J, Liu M, Deng S, Chen P. Polymorphisms of Base-Excision Repair Genes hOGG1 326cys and XRCC1 280His Increase Hepatocellular Carcinoma Risk. Digestive Diseases and Sciences. 2012;57(9):2451-7.
82. Pan HZ, Liang J, Yu Z, Lun LM, Li H, Wang Q. Polymorphism of DNA repair gene XRCC1 and hepatocellular carcinoma risk in Chinese population. Asian Pacific journal of cancer prevention : APJCP. 2011;12(11):2947-50.
83. Guo LY, Jin XP, Niu W, Li XF, Liu BH, Wang YL. Association of XPD and XRCC1 genetic polymorphisms with hepatocellular carcinoma risk. Asian Pacific journal of cancer prevention : APJCP. 2012;13(9):4423-6.
84. Chen C-C, Yang S-Y, Liu C-J, Lin C-L, Liaw Y-F, Lin S-M, et al. Association of cytokine and DNA repair gene polymorphisms with hepatitis B-related hepatocellular carcinoma. International Journal of Epidemiology. 2005;34(6):1310-8.
85. Mattar MM, Zekri AN, Hussein N, Morsy H, Esmat G, Amin MA. Polymorphisms of base-excision repair genes and the hepatocarcinogenesis. Gene. 2018;675:62-8.
86. Long XD, Ma Y, Zhou YF, Yao JG, Ban FZ, Huang YZ, et al. XPD codon 312 and 751 polymorphisms, and AFB1 exposure, and hepatocellular carcinoma risk. BMC cancer. 2009;9:400.
87. Bayram S, Akkız H, Bekar A, Akgöllü E, Yıldırım S. The significance of Exonuclease 1 K589E polymorphism on hepatocellular carcinoma susceptibility in the Turkish population: a case–control study. Molecular Biology Reports. 2012;39(5):5943-51.
88. Huang MD, Chen XF, Xu G, Wu QQ, Zhang JH, Chen GF, et al. Genetic variation in the NBS1 gene is associated with hepatic cancer risk in a Chinese population. DNA and cell biology. 2012;31(5):678-82.
89. Zhen Y, Xiao R, Chen X, Yuan C, Sun Y, Li J. A non-synonymous polymorphism in NBS1 is associated with progression from chronic hepatitis B virus infection to hepatocellular carcinoma in a Chinese population. OncoTargets and therapy. 2018;11:563-9.
90. Li Z, Guo Y, Zhou L, Ge Y, Wei L, Li L, et al. Association of a functional RAD52 genetic variant locating in a miRNA binding site with risk of HBV-related hepatocellular carcinoma. Molecular Carcinogenesis. 2015;54(9):853-8.
91. Hsieh Y-H, Chang W-S, Tsai C-W, Tsai J-P, Hsu C-M, Jeng L-B, et al. DNA double-strand break repair gene XRCC7 genotypes were associated with hepatocellular carcinoma risk in Taiwanese males and alcohol drinkers. Tumor Biology. 2015;36(6):4101-6.
92. Zhu X, Wang Z, Qiu X, Wang W, Bei C, Tan C, et al. Rs2303428 of MSH2 Is Associated with Hepatocellular Carcinoma Prognosis in a Chinese Population. DNA and cell biology. 2018;37(7):634-41.
93. Zhu X, Liu W, Qiu X, Wang Z, Tan C, Bei C, et al. Single nucleotide polymorphisms in MLH1 predict poor prognosis of hepatocellular carcinoma in a Chinese population. Oncotarget. 2017;8(45):80039-49.
94. Spivak G. Nucleotide excision repair in humans. DNA repair. 2015;36:13-8.
95. Vaisman A, Woodgate R. Translesion DNA polymerases in eukaryotes: what makes them tick? Critical Reviews in Biochemistry and Molecular Biology. 2017;52(3):274-303.
96. Prakash S, Johnson RE, Prakash L. EUKARYOTIC TRANSLESION SYNTHESIS DNA POLYMERASES: Specificity of Structure and Function. Annual Review of Biochemistry. 2005;74(1):317-53.
97. Guengerich FP, Zhao L, Pence MG, Egli M. Structure and function of the translesion DNA polymerases and interactions with damaged DNA. Perspectives in Science. 2015;4:24-31.
98. Hedglin M, Benkovic SJ. Regulation of Rad6/Rad18 Activity During DNA Damage Tolerance. Annual Review of Biophysics. 2015;44(1):207-28.
99. Bedard LL, Massey TE. Aflatoxin B1-induced DNA damage and its repair. Cancer Letters. 2006;241(2):174-83.
100. Leadon SA, Tyrrell RM, Cerutti PA. Excision Repair of Aflatoxin B₁-DNA Adducts in Human Fibroblasts. Cancer Research. 1981;41(12 Part 1):5125.
101. Takahashi Y, Nakatsuru Y, Zhang S, Shimizu Y, Kume H, Tanaka K, et al. Enhanced spontaneous and aflatoxin-induced liver tumorigenesis in xeroderma pigmentosum group A gene-deficient mice. Carcinogenesis. 2002;23(4):627-33.

102. Hsu IC, Metcalf RA, Sun T, Welsh JA, Wang NJ, Harris CC. Mutational hot spot in the p53 gene in human hepatocellular carcinomas. Nature. 1991;350(6317):427-8.
103. Bressac B, Kew M, Wands J, Ozturk M. Selective G to T mutations of p53 gene in hepatocellular carcinoma from southern Africa. Nature. 1991;350(6317):429-31.
104. Besaratinia A, Kim S-i, Hainaut P, Pfeifer GP. In Vitro Recapitulating of TP53 Mutagenesis in Hepatocellular Carcinoma Associated With Dietary Aflatoxin B1 Exposure. Gastroenterology. 2009;137(3):1127-37.e5.
105. Lin Y-C, Li L, Makarova AV, Burgers PM, Stone MP, Lloyd RS. Molecular basis of aflatoxin-induced mutagenesis—role of the aflatoxin B1-formamidopyrimidine adduct. Carcinogenesis. 2014;35(7):1461-8.
106. Lin YC, Owen N, Minko IG, Lange SS, Tomida J, Li L, et al. DNA polymerase zeta limits chromosomal damage and promotes cell survival following aflatoxin exposure. Proceedings of the National Academy of Sciences of the United States of America. 2016;113(48):13774-9.
107. Abdel Ghaffar TY, Elsobky ES, Elsayed SM. Cholestasis in patients with Cockayne syndrome and suggested modified criteria for clinical diagnosis. Orphanet journal of rare diseases. 2011;6:13-.
108. Uehara Y, Ikehata H, Furuya M, Kobayashi S, He D, Chen Y, et al. XPC is involved in genome maintenance through multiple pathways in different tissues. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2009;670(1):24-31.
109. Fautrel A, Andrieux L, Musso O, Boudjema K, Guillouzo A, Langouët S. Overexpression of the two nucleotide excision repair genes ERCC1 and XPC in human hepatocellular carcinoma. Journal of Hepatology. 2005;43(2):288-93.
110. Shi WK, Zhu XD, Wang CH, Zhang YY, Cai H, Li XL, et al. PFKFB3 blockade inhibits hepatocellular carcinoma growth by impairing DNA repair through AKT. Cell death & disease. 2018;9(4):428.
111. Peng Q, Li S, Lao X, Chen Z, Li R, Qin X. Association between XPD Lys751Gln and Asp312Asn polymorphisms and hepatocellular carcinoma risk: a systematic review and meta-analysis. Medicine. 2014;93(29):e330.
112. Shi Y-H, Wang B, Xu B-P, Jiang D-N, Zhao D-M, Ji M-R, et al. The association of six non-synonymous variants in three DNA repair genes with hepatocellular carcinoma risk: a meta-analysis. Journal of cellular and molecular medicine. 2016;20(11):2056-63.
113. Blackford AN, Jackson SP. ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. Molecular cell. 2017;66(6):801-17.
114. Marechal A, Zou L. DNA damage sensing by the ATM and ATR kinases. Cold Spring Harbor perspectives in biology. 2013;5(9).
115. Chatterjee N, Walker GC. Mechanisms of DNA damage, repair, and mutagenesis. Environmental and molecular mutagenesis. 2017;58(5):235-63.
116. Lieber MR. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. Annual Review of Biochemistry. 2010;79(1):181-211.
117. Ramsden DA. Polymerases in nonhomologous end joining: building a bridge over broken chromosomes. Antioxid Redox Signal. 2011;14(12):2509-19.
118. Li X, Heyer W-D. Homologous recombination in DNA repair and DNA damage tolerance. Cell research. 2008;18(1):99-113.
119. Moynahan ME, Jasin M. Loss of heterozygosity induced by a chromosomal double-strand break. Proceedings of the National Academy of Sciences of the United States of America. 1997;94(17):8988-93.
120. Mehta A, Haber JE. Sources of DNA double-strand breaks and models of recombinational DNA repair. Cold Spring Harbor perspectives in biology. 2014;6(9):a016428.
121. Grant DM. Detoxification pathways in the liver. J Inherit Metab Dis. 1991;14(4):421-30.
122. Letouzé E, Shinde J, Renault V, Couchy G, Blanc J-F, Tubacher E, et al. Mutational signatures reveal the dynamic interplay of risk factors and cellular processes during liver tumorigenesis. Nature communications. 2017;8(1):1315.
123. Wang Y, Li M, Long J, Shi X-Y, Li Q, Chen J, et al. Clinical significance of increased expression of Nijmegen breakage syndrome gene (NBS1) in human primary liver cancer. Hepatology International. 2014;8(2):250-9.

124. Evert M, Frau M, Tomasi ML, Latte G, Simile MM, Seddaiu MA, et al. Deregulation of DNA-dependent protein kinase catalytic subunit contributes to human hepatocarcinogenesis development and has a putative prognostic value. British journal of cancer. 2013;109:2654.
125. Dai Y, Tang Z, Yang Z, Zhang L, Deng Q, Zhang X, et al. EXO1 overexpression is associated with poor prognosis of hepatocellular carcinoma patients. Cell cycle (Georgetown, Tex). 2018;17(19-20):2386-97.
126. Lin J, Shi J, Guo H, Yang X, Jiang Y, Long J, et al. Alterations in DNA damage repair genes in primary liver cancer. Clinical Cancer Research. 2019:clincanres.0127.2019.
127. Dumon-Jones V, Frappart PO, Tong WM, Sajithlal G, Hulla W, Schmid G, et al. Nbn heterozygosity renders mice susceptible to tumor formation and ionizing radiation-induced tumorigenesis. Cancer Res. 2003;63(21):7263-9.
128. Kurimasa A, Burma S, Henrie M, Ouyang H, Osaki M, Ito H, et al. Disruption of NBS1 gene leads to early embryonic lethality in homozygous null mice and induces specific cancer in heterozygous mice. United States; 2002 2002-04-15.
129. Wang Y, Hong Y, Li M, Long J, Zhao Y-P, Zhang J-X, et al. Mutation inactivation of Nijmegen breakage syndrome gene (NBS1) in hepatocellular carcinoma and intrahepatic cholangiocarcinoma. PloS one. 2013;8(12):e82426-e.
130. Teoh NC, Dan YY, Swisshelm K, Lehman S, Wright JH, Haque J, et al. Defective DNA strand break repair causes chromosomal instability and accelerates liver carcinogenesis in mice. Hepatology (Baltimore, Md). 2008;47(6):2078-88.
131. Ge C, Vilfranc CL, Che L, Pandita RK, Hambarde S, Andreassen PR, et al. The BRUCE-ATR signaling axis is required for accurate DNA replication and suppression of liver cancer development. Hepatology (Baltimore, Md). 2019.
132. Kawai H, Suda T, Aoyagi Y, Isokawa O, Mita Y, Waguri N, et al. Quantitative evaluation of genomic instability as a possible predictor for development of hepatocellular carcinoma: Comparison of loss of heterozygosity and replication error. Hepatology (Baltimore, Md). 2000;31(6):1246-50.
133. Machida K, McNamara G, Cheng KTH, Huang J, Wang C-H, Comai L, et al. Hepatitis C Virus Inhibits DNA Damage Repair through Reactive Oxygen and Nitrogen Species and by Interfering with the ATM-NBS1/Mre11/Rad50 DNA Repair Pathway in Monocytes and Hepatocytes. The Journal of Immunology. 2010;185(11):6985.
134. Carson CT, Schwartz RA, Stracker TH, Lilley CE, Lee DV, Weitzman MD. The Mre11 complex is required for ATM activation and the G₂/M checkpoint. The EMBO Journal. 2003;22(24):6610.
135. Nguyen TTT, Park EM, Lim YS, Hwang SB. Nonstructural Protein 5A Impairs DNA Damage Repair: Implications for Hepatitis C Virus-Mediated Hepatocarcinogenesis. Journal of virology. 2018;92(11).
136. Ko H-L, Ren E-C. Novel poly (ADP-ribose) polymerase 1 binding motif in hepatitis B virus core promoter impairs DNA damage repair. Hepatology (Baltimore, Md). 2011;54(4):1190-8.
137. Pham HT, Nguyen TTT, Nguyen LP, Han SS, Lim YS, Hwang SB. Hepatitis C Virus Downregulates Ubiquitin-Conjugating Enzyme E2S Expression To Prevent Proteasomal Degradation of NS5A, Leading to Host Cells More Sensitive to DNA Damage. Journal of virology. 2019;93(2).
138. Li G-M. Mechanisms and functions of DNA mismatch repair. Cell Research. 2007;18:85.
139. Liu D, Keijzers G, Rasmussen LJ. DNA mismatch repair and its many roles in eukaryotic cells. Mutation Research/Reviews in Mutation Research. 2017;773:174-87.
140. Sharma S, Salehi F, Scheithauer BW, Rotondo F, Syro LV, Kovacs K. Role of MGMT in tumor development, progression, diagnosis, treatment and prognosis. Anticancer Res. 2009;29(10):3759-68.
141. Mao SA, Glorioso JM, Nyberg SL. Liver regeneration. Translational research : the journal of laboratory and clinical medicine. 2014;163(4):352-62.
142. de la Chapelle A. Microsatellite instability. The New England journal of medicine. 2003;349(3):209-10.
143. Takagi K, Esumi M, Takano S, Iwai S. Replication error frequencies in primary hepatocellular carcinoma: a comparison of solitary primary versus multiple primary cancers. Liver. 1998;18(4):272-6.

144. Wang L, Bani-Hani A, Montoya DP, Roche PC, Thibodeau SN, Burgart LJ, et al. hMLH1 and hMSH2 expression in human hepatocellular carcinoma. International journal of oncology. 2001;19(3):567-70.
145. Kim ST, Klempner SJ, Park SH, Park JO, Park YS, Lim HY, et al. Correlating programmed death ligand 1 (PD-L1) expression, mismatch repair deficiency, and outcomes across tumor types: implications for immunotherapy. Oncotarget. 2017;8(44):77415-23.
146. Togni R, Bagla N, Muiesan P, Miquel R, O'Grady J, Heaton N, et al. Microsatellite instability in hepatocellular carcinoma in non-cirrhotic liver in patients older than 60 years. Hepatology Research. 2009;39(3):266-73.
147. Casper M, Weber SN, Kloor M, Müllenbach R, Grobholz R, Lammert F, et al. Hepatocellular carcinoma as extracolonic manifestation of Lynch syndrome indicates SEC63 as potential target gene in hepatocarcinogenesis. Scandinavian Journal of Gastroenterology. 2013;48(3):344-51.
148. Zekri AR, Sabry GM, Bahnassy AA, Shalaby KA, Abdel-Wahabh SA, Zakaria S. Mismatch repair genes (hMLH1, hPMS1, hPMS2, GTBP/hMSH6, hMSH2) in the pathogenesis of hepatocellular carcinoma. World J Gastroenterol. 2005;11(20):3020-6.
149. Helal TE, Khamis NS, El-Sharkawy TM, Nada OH, Radwan NA. Immunohistochemical expression of mismatch repair genes (hMSH2 and hMLH1) in hepatocellular carcinoma in Egypt. APMIS. 2010;118(12):934-40.
150. Wani Y, Notohara K, Tsukayama C, Okada S. Reduced expression of hMLH1 and hMSH2 gene products in high-grade hepatocellular carcinoma. Acta medica Okayama. 2001;55(2):65-71.
151. Wong YH, Wu CC, Lin CL, Chen TS, Chang TH, Chen BS. Applying NGS Data to Find Evolutionary Network Biomarkers from the Early and Late Stages of Hepatocellular Carcinoma. BioMed research international. 2015;2015:391475.
152. Schmutte C, Sadoff MM, Shim K-S, Acharya S, Fishel R. The Interaction of DNA Mismatch Repair Proteins with Human Exonuclease I. Journal of Biological Chemistry. 2001;276(35):33011-8.
153. Shiraha H, Yamamoto K, Namba M. Human hepatocyte carcinogenesis (review). International journal of oncology. 2013;42(4):1133-8.
154. Williams AB, Schumacher B. p53 in the DNA-Damage-Repair Process. Cold Spring Harb Perspect Med. 2016;6(5):10.1101/cshperspect.a026070 a.
155. Macdonald GA, Greenson JK, Saito K, Cherian SP, Appelman HD, Boland CR. Microsatellite instability and loss of heterozygosity at DNA mismatch repair gene loci occurs during hepatic carcinogenesis. Hepatology (Baltimore, Md). 1998;28(1):90-7.
156. Matsukura S, Soejima H, Nakagawachi T, Yakushiji H, Ogawa A, Fukuhara M, et al. CpG methylation of MGMT and hMLH1 promoter in hepatocellular carcinoma associated with hepatitis viral infection. British journal of cancer. 2003;88(4):521-9.
157. Zekri Ael R, Nassar AA, El-Din El-Rouby MN, Shousha HI, Barakat AB, El-Desouky ED, et al. Disease progression from chronic hepatitis C to cirrhosis and hepatocellular carcinoma is associated with increasing DNA promoter methylation. Asian Pacific journal of cancer prevention : APJCP. 2014;14(11):6721-6.
158. Le Grazie M, Biagini MR, Tarocchi M, Polvani S, Galli A. Chemotherapy for hepatocellular carcinoma: The present and the future. World journal of hepatology. 2017;9(21):907-20.
159. Marin JJG, Macias RIR, Monte MJ, Romero MR, Asensio M, Sanchez-Martin A, et al. Molecular Bases of Drug Resistance in Hepatocellular Carcinoma. Cancers. 2020;12(6).
160. Kudo M, Finn RS, Qin S, Han KH, Ikeda K, Piscaglia F, et al. Lenvatinib versus sorafenib in first-line treatment of patients with unresectable hepatocellular carcinoma: a randomised phase 3 non-inferiority trial. Lancet. 2018;391(10126):1163-73.
161. Chae YK, Anker JF, Carneiro BA, Chandra S, Kaplan J, Kalyan A, et al. Genomic landscape of DNA repair genes in cancer. Oncotarget. 2016;7(17):23312-21.
162. Nuthalapati S, Munasinghe W, Giranda V, Xiong H. Clinical Pharmacokinetics and Mass Balance of Veliparib in Combination with Temozolomide in Subjects with Nonhematologic Malignancies. Clin Pharmacokinet. 2018;57(1):51-8.
163. Gabrielson A, Tesfaye AA, Marshall JL, Pishvaian MJ, Smaglo B, Jha R, et al. Phase II study of temozolomide and veliparib combination therapy for sorafenib-refractory advanced hepatocellular carcinoma. Cancer Chemother Pharmacol. 2015;76(5):1073-9.

164. Huang CY, Hsieh FS, Wang CY, Chen LJ, Chang SS, Tsai MH, et al. Palbociclib enhances radiosensitivity of hepatocellular carcinoma and cholangiocarcinoma via inhibiting ataxia telangiectasia-mutated kinase-mediated DNA damage response. European journal of cancer (Oxford, England : 1990). 2018;102:10-22.
165. Fang F, Munck J, Tang J, Taverna P, Wang Y, Miller DF, et al. The novel, small-molecule DNA methylation inhibitor SGI-110 as an ovarian cancer chemosensitizer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2014;20(24):6504-16.
166. Adamsen BL, Kravik KL, De Angelis PM. DNA damage signaling in response to 5-fluorouracil in three colorectal cancer cell lines with different mismatch repair and TP53 status. International journal of oncology. 2011;39(3):673-82.
167. Leemans CR, Snijders PJF, Brakenhoff RH. The molecular landscape of head and neck cancer. Nature reviews Cancer. 2018;18(5):269-82.
168. Elwood JM, Youlden DR, Chelimo C, Ioannides SJ, Baade PD. Comparison of oropharyngeal and oral cavity squamous cell cancer incidence and trends in New Zealand and Queensland, Australia. Cancer epidemiology. 2014;38(1):16-21.
169. Australian Bureau of Statistics. Report 4364.0.55.001. 2015.
170. Mellin H, Friesland S, Lewensohn R, Dalianis T, Munck-Wikland E. Human papillomavirus (HPV) DNA in tonsillar cancer: clinical correlates, risk of relapse, and survival. Int J Cancer. 2000;89(3):300-4.
171. Chen AM, Chen LM, Vaughan A, Sreeraman R, Farwell DG, Luu Q, et al. Tobacco smoking during radiation therapy for head-and-neck cancer is associated with unfavorable outcome. International journal of radiation oncology, biology, physics. 2011;79(2):414-9.
172. Vlashi E, Chen AM, Boyrie S, Yu G, Nguyen A, Brower PA, et al. Radiation-Induced Dedifferentiation of Head and Neck Cancer Cells Into Cancer Stem Cells Depends on Human Papillomavirus Status. International journal of radiation oncology, biology, physics. 2016;94(5):1198-206.
173. Kalavrezos N, Bhandari R. Current trends and future perspectives in the surgical management of oral cancer. Oral Oncol. 2010;46(6):429-32.
174. Reid P, Wilson P, Li Y, Marcu LG, Staudacher AH, Brown MP, et al. In vitro investigation of head and neck cancer stem cell proportions and their changes following X-ray irradiation as a function of HPV status. PloS one. 2017;12(10):e0186186.
175. Vilalta M, Rafat M, Graves EE. Effects of radiation on metastasis and tumor cell migration. Cellular and molecular life sciences : CMLS. 2016;73(16):2999-3007.
176. Strong MS, Vaughan CW, Kayne HL, Aral IM, Ucmakli A, Feldman M, et al. A randomized trial of preoperative radiotherapy in cancer of the oropharynx and hypopharynx. American journal of surgery. 1978;136(4):494-500.
177. von Essen CF. Radiation enhancement of metastasis: a review. Clinical & experimental metastasis. 1991;9(2):77-104.
178. Kumar KV, Hema KN. Extracellular matrix in invasion and metastasis of oral squamous cell carcinoma. J Oral Maxillofac Pathol. 2019;23(1):10-6.
179. Thomas GJ, Speight PM. Cell adhesion molecules and oral cancer. Crit Rev Oral Biol Med. 2001;12(6):479-98.
180. Parri M, Chiarugi P. Rac and Rho GTPases in cancer cell motility control. Cell Commun Signal. 2010;8:23.
181. Chiang SP, Cabrera RM, Segall JE. Tumor cell intravasation. Am J Physiol Cell Physiol. 2016;311(1):C1-C14.
182. Tan GJ, Peng ZK, Lu JP, Tang FQ. Cathepsins mediate tumor metastasis. World J Biol Chem. 2013;4(4):91-101.
183. Prasad S, Ramachandran S, Gupta N, Kaushik I, Srivastava SK. Cancer cells stemness: A doorstep to targeted therapy. Biochimica et biophysica acta Molecular basis of disease. 2020;1866(4):165424.
184. Lim YC, Oh SY, Cha YY, Kim SH, Jin X, Kim H. Cancer stem cell traits in squamospheres derived from primary head and neck squamous cell carcinomas. Oral Oncol. 2011;47(2):83-91.

185. Kaseb HO, Fohrer-Ting H, Lewis DW, Lagasse E, Gollin SM. Identification, expansion and characterization of cancer cells with stem cell properties from head and neck squamous cell carcinomas. Exp Cell Res. 2016;348(1):75-86.
186. Reid PA, Wilson P, Li Y, Marcu LG, Bezak E. Current understanding of cancer stem cells: Review of their radiobiology and role in head and neck cancers. Head & neck. 2017;39(9):1920-32.
187. Chen C, Zhao S, Karnad A, Freeman JW. The biology and role of CD44 in cancer progression: therapeutic implications. J Hematol Oncol. 2018;11(1):64.
188. Clark DW, Palle K. Aldehyde dehydrogenases in cancer stem cells: potential as therapeutic targets. Annals of translational medicine. 2016;4(24):518.
189. An Y, Kiang A, Lopez JP, Kuo SZ, Yu MA, Abhold EL, et al. Cigarette smoke promotes drug resistance and expansion of cancer stem cell-like side population. PloS one. 2012;7(11):e47919.
190. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):550.
191. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4(1521):1521.
192. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. Journal of Open Source Software. 2019;4(43):1686.
193. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model: Springer New York; 2013.
194. Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, et al. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. Nucleic Acids Res. 2014;42(17):e133.
195. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010 [Available from: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
196. Ewels P, Magnusson M, Lundin S, Kaller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics (Oxford, England). 2016;32(19):3047-8.
197. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC bioinformatics. 2011;12(1):323.
198. Gao C-H. ggVennDiagram: A 'ggplot2' Implement of Venn Diagram. 0.5.0. ed2021.
199. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC bioinformatics. 2009;10(1):48.
200. Farré D, Roset R, Huerta M, Adsuara JE, Roselló L, Albà MM, et al. Identification of patterns in biological sequences at the ALGGEN server: PROMO and MALGEN. Nucleic acids research. 2003;31(13):3651-3.
201. Messeguer X, Escudero R, Farré D, Núñez O, Martínez J, Albà MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinformatics (Oxford, England). 2002;18(2):333-4.
202. Mosca R, Ceol A, Aloy P. Interactome3D: adding structural details to protein networks. Nat Methods. 2013;10(1):47-53.
203. Chen YA, Tripathi LP, Fujiwara T, Kameyama T, Itoh MN, Mizuguchi K. The TargetMine Data Warehouse: Enhancement and Updates. Front Genet. 2019;10(934):934.
204. Oughtred R, Rust J, Chang C, Breitkreutz BJ, Stark C, Willems A, et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. Protein science : a publication of the Protein Society. 2021;30(1):187-200.
205. Bernhofer M, Goldberg T, Wolf S, Ahmed M, Zaugg J, Boden M, et al. NLSdb-major update for database of nuclear localization signals and nuclear export signals. Nucleic Acids Res. 2018;46(D1):D503-D8.
206. Kosugi S, Hasebe M, Tomita M, Yanagawa H. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(25):10171-6.
207. Kumar M, Gouw M, Michael S, Samano-Sanchez H, Pancsa R, Glavina J, et al. ELM-the eukaryotic linear motif resource in 2020. Nucleic Acids Res. 2020;48(D1):D296-D306.
208. Yang Y, Ding X, Zhu G, Niroula A, Lv Q, Vihinen M. ProTstab - predictor for cellular protein stability. BMC Genomics. 2019;20(1):804.

209. Yachdav G, Kloppmann E, Kajan L, Hecht M, Goldberg T, Hamp T, et al. PredictProtein--an open resource for online prediction of protein structural and functional features. Nucleic Acids Res. 2014;42(Web Server issue):W337-43.
210. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. PloS one. 2011;6(7):e21800.
211. Franken NA, Rodermond HM, Stap J, Haveman J, van Bree C. Clonogenic assay of cells in vitro. Nat Protoc. 2006;1(5):2315-9.
212. Lu Y, Liu Y, Yang C. Evaluating In Vitro DNA Damage Using Comet Assay. Journal of visualized experiments : JoVE. 2017(128).
213. Olive PL, Banath JP. The comet assay: a method to measure DNA damage in individual cells. Nat Protoc. 2006;1(1):23-9.
214. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 2018;46(W1):W296-w303.
215. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Res. 2005;33(Web Server issue):W89-93.
216. Wu EM, Wong LL, Hernandez BY, Ji JF, Jia W, Kwee SA, et al. Gender differences in hepatocellular cancer: disparities in nonalcoholic fatty liver disease/steatohepatitis and liver transplantation. Hepatoma Res. 2018;4:66.
217. Gong F, Miller KM. Histone methylation and the DNA damage response. Mutat Res. 2019;780:37-47.
218. Arguello G, Balboa E, Arrese M, Zanlungo S. Recent insights on the role of cholesterol in non-alcoholic fatty liver disease. Biochim Biophys Acta. 2015;1852(9):1765-78.
219. Kneeman JM, Misdraji J, Corey KE. Secondary causes of nonalcoholic fatty liver disease. Therap Adv Gastroenterol. 2012;5(3):199-207.
220. Lin HY, Lai RH, Lin ST, Lin RC, Wang MJ, Lin CC, et al. Suppressor of cytokine signaling 6 (SOCS6) promotes mitochondrial fission via regulating DRP1 translocation. Cell Death Differ. 2013;20(1):139-53.
221. Carbone M, Yang H, Pass HI, Krausz T, Testa JR, Gaudino G. BAP1 and cancer. Nature reviews Cancer. 2013;13(3):153-9.
222. Lu Y, Adegoke OA, Nepveu A, Nakayama KI, Bedard N, Cheng D, et al. USP19 deubiquitinating enzyme supports cell proliferation by stabilizing KPC1, a ubiquitin ligase for p27Kip1. Mol Cell Biol. 2009;29(2):547-58.
223. Wang OH, Azizian N, Guo M, Capello M, Deng D, Zang F, et al. Prognostic and Functional Significance of MAP4K5 in Pancreatic Cancer. PloS one. 2016;11(3):e0152300.
224. Wang S, Xu M, Sun Z, Yu X, Deng Y, Chang H. LINC01018 confers a novel tumor suppressor role in hepatocellular carcinoma through sponging microRNA-182-5p. American journal of physiology Gastrointestinal and liver physiology. 2019;317(2):G116-G26.
225. Cao MQ, You AB, Zhu XD, Zhang W, Zhang YY, Zhang SZ, et al. miR-182-5p promotes hepatocellular carcinoma progression by repressing FOXO3a. J Hematol Oncol. 2018;11(1):12.
226. Liu W, Li X, Chu ES, Go MY, Xu L, Zhao G, et al. Paired box gene 5 is a novel tumor suppressor in hepatocellular carcinoma through interaction with p53 signaling pathway. Hepatology (Baltimore, Md). 2011;53(3):843-53.
227. Gorla-Bajszczak A, Juge-Aubry C, Pernin A, Burger AG, Meier CA. Conserved amino acids in the ligand-binding and tau(i) domains of the peroxisome proliferator-activated receptor alpha are necessary for heterodimerization with RXR. Molecular and cellular endocrinology. 1999;147(1-2):37-47.
228. Massip-Salcedo M, Zaouali MA, Padrissa-Altes S, Casillas-Ramirez A, Rodes J, Rosello-Catafau J, et al. Activation of peroxisome proliferator-activated receptor-alpha inhibits the injurious effects of adiponectin in rat steatotic liver undergoing ischemia-reperfusion. Hepatology (Baltimore, Md). 2008;47(2):461-72.
229. Chen R, Chen Y, Zhao W, Fang C, Zhou W, Yang X, et al. The Role of Methyltransferase NSD2 as a Potential Oncogene in Human Solid Tumors. OncoTargets and therapy. 2020;13:6837-46.
230. Zhang J, Lee YR, Dang F, Gan W, Menon AV, Katon JM, et al. PTEN Methylation by NSD2 Controls Cellular Sensitivity to DNA Damage. Cancer Discov. 2019;9(9):1306-23.

231. Shi L, Feng Y, Lin H, Ma R, Cai X. Role of estrogen in hepatocellular carcinoma: is inflammation the key? J Transl Med. 2014;12:93.
232. Wu Q, Jubb H, Blundell TL. Phosphopeptide interactions with BRCA1 BRCT domains: More than just a motif. Prog Biophys Mol Biol. 2015;117(2-3):143-8.
233. Wei Y, Wang D, Moran G, Estrada A, Pagliassotti MJ. Fructose-induced stress signaling in the liver involves methylglyoxal. Nutr Metab (Lond). 2013;10(1):32.
234. Mehruba M, Siddique SM, Mukai H. PKN1 controls the aggregation, spheroid formation, and viability of mouse embryonic fibroblasts in suspension culture. Biochemical and biophysical research communications. 2020;523(2):398-404.
235. Kubouchi K, Mukai H. PKN2 is involved in aggregation and spheroid formation of fibroblasts in suspension culture by regulating cell motility and N-cadherin expression. Biochemistry and biophysics reports. 2021;25:100895.
236. Kapalczynska M, Kolenda T, Przybyla W, Zajaczkowska M, Teresiak A, Filas V, et al. 2D and 3D cell cultures - a comparison of different types of cancer cell cultures. Arch Med Sci. 2018;14(4):910-9.
237. Johnson S, Chen H, Lo PK. In vitro Tumorsphere Formation Assays. Bio Protoc. 2013;3(3):e325.
238. Kim H, Lin Q, Yun Z. BRCA1 regulates the cancer stem cell fate of breast cancer cells in the context of hypoxia and histone deacetylase inhibitors. Scientific reports. 2019;9(1):9702.
239. Bai F, Chan HL, Scott A, Smith MD, Fan C, Herschkowitz JI, et al. BRCA1 suppresses epithelial-to-mesenchymal transition and stem cell dedifferentiation during mammary and tumor development. Cancer Res. 2014;74(21):6161-72.
240. Fontana A, de Laureto PP, Spolaore B, Frare E. Identifying disordered regions in proteins by limited proteolysis. Methods in molecular biology (Clifton, NJ). 2012;896:297-318.
241. Uversky VN. Intrinsically Disordered Proteins and Their "Mysterious" (Meta)Physics. Frontiers in Physics. 2019;7(10).
242. Zeisberg M, Neilson EG. Biomarkers for epithelial-mesenchymal transitions. The Journal of clinical investigation. 2009;119(6):1429-37.
243. Cortesina G, Martone T. Molecular metastases markers in head and neck squamous cell carcinoma: review of the literature. Acta otorhinolaryngologica Italica : organo ufficiale della Societa italiana di otorinolaringologia e chirurgia cervico-facciale. 2006;26(6):317-25.
244. Zhao W, Li Y, Zhang X. Stemness-Related Markers in Cancer. Cancer translational medicine. 2017;3(3):87-95.
245. Pardo-Reoyo S, Roig-Lopez JL, Yang ES. Potential biomarkers for radiosensitivity in head and neck cancers. Annals of translational medicine. 2016;4(24):524.
246. Loizzi V, Del Vecchio V, Gargano G, De Liso M, Kardashi A, Naglieri E, et al. Biological Pathways Involved in Tumor Angiogenesis and Bevacizumab Based Anti-Angiogenic Therapy with Special References to Ovarian Cancer. International journal of molecular sciences. 2017;18(9).
247. Dasgupta P, Rizwani W, Pillai S, Kinkade R, Kovacs M, Rastogi S, et al. Nicotine induces cell proliferation, invasion and epithelial-mesenchymal transition in a variety of human cancer cell lines. Int J Cancer. 2009;124(1):36-45.
248. Deng Y, Wan Q, Yan W. Integrin alpha5/ITGA5 Promotes The Proliferation, Migration, Invasion And Progression Of Oral Squamous Carcinoma By Epithelial-Mesenchymal Transition. Cancer Manag Res. 2019;11:9609-20.
249. Singh B, Berry JA, Shoher A, Ramakrishnan V, Lucci A. COX-2 overexpression increases motility and invasion of breast cancer cells. International journal of oncology. 2005;26(5):1393-9.
250. Munn LL, Garkavtsev I. SLPI: a new target for stopping metastasis. Aging. 2018;10(1):13-4.
251. Jedroszka D, Orzechowska M, Hamouz R, Gorniak K, Bednarek AK. Markers of epithelial-to-mesenchymal transition reflect tumor biology according to patient age and Gleason score in prostate cancer. PloS one. 2017;12(12):e0188842.
252. Martuszewska D, Ljungberg B, Johansson M, Landberg G, Oslakovic C, Dahlback B, et al. Tensin3 is a negative regulator of cell migration and all four Tensin family members are downregulated in human kidney cancer. PloS one. 2009;4(2):e4350.
253. Sarwar R, Pierce N, Koppe S. Obesity and nonalcoholic fatty liver disease: current perspectives. Diabetes Metab Syndr Obes. 2018;11:533-42.

254. Basciano H, Federico L, Adeli K. Fructose, insulin resistance, and metabolic dyslipidemia. Nutr Metab (Lond). 2005;2(1):5.
255. Gillman R, Lopes Floro K, Wankell M, Hebbard L. The role of DNA damage and repair in liver cancer. Biochimica et biophysica acta Reviews on cancer. 2021;1875(1):188493.
256. Wolters JEJ, van Breda SGJ, Grossmann J, Fortes C, Caiment F, Kleinjans JCS. Integrated 'omics analysis reveals new drug-induced mitochondrial perturbations in human hepatocytes. Toxicology letters. 2018;289:1-13.
257. Tong V, Teng XW, Chang TK, Abbott FS. Valproic acid II: effects on oxidative stress, mitochondrial membrane potential, and cytotoxicity in glutathione-depleted rat hepatocytes. Toxicological sciences : an official journal of the Society of Toxicology. 2005;86(2):436-43.
258. Wolters J, van Breda S, Claessen S, de Kok T, Kleinjans J. Data on novel DNA methylation changes induced by valproic acid in human hepatocytes. Data in brief. 2018;16:161-71.
259. Mao Z, Bozzella M, Seluanov A, Gorbunova V. Comparison of nonhomologous end joining and homologous recombination in human cells. DNA repair. 2008;7(10):1765-71.
260. Boucher J, Kleinridders A, Kahn CR. Insulin receptor signaling in normal and insulin-resistant states. Cold Spring Harbor perspectives in biology. 2014;6(1):a009191.
261. Choi E, Zhang X, Xing C, Yu H. Mitotic Checkpoint Regulators Control Insulin Signaling and Metabolic Homeostasis. Cell. 2016;166(3):567-81.
262. Ridley AJ. RhoA, RhoB and RhoC have different roles in cancer cell migration. J Microsc. 2013;251(3):242-9.
263. Chang PM, Chen CH, Yeh CC, Lu HJ, Liu TT, Chen MH, et al. Transcriptome analysis and prognosis of ALDH isoforms in human cancer. Scientific reports. 2018;8(1):2713.
264. Ostroff JS, Jacobsen PB, Moadel AB, Spiro RH, Shah JP, Strong EW, et al. Prevalence and predictors of continued tobacco use after treatment of patients with head and neck cancer. Cancer. 1995;75(2):569-76.

Appendix

Appendix A: Top 50 genes significantly mutated in High-HORMAD2 patients. LH, Low HORMAD2; NH, Normal HORMAD2; HH, High HORMAD2; OR, Odd's Ratio.

	Mutation Frequency			Fisher's Exact Testing	
Gene	**LH (n=92)**	**NH (n=237)**	**HH (n=38)**	**p**	**OR (HH vs NH)**
PRKX	0.00%	0.00%	13.16%	3.97E-05	Inf
ACVR2A	1.09%	2.53%	18.42%	5.21E-04	8.574425
STAG1	1.09%	0.42%	10.53%	0.001424	27.14709
G6PC	0.00%	0.00%	7.89%	0.002461	Inf
CAPRIN2	0.00%	0.00%	7.89%	0.002461	Inf
NFKB1	0.00%	0.00%	7.89%	0.002461	Inf
PER3	2.17%	0.00%	7.89%	0.002461	Inf
CES1	0.00%	0.00%	7.89%	0.002461	Inf
CASP3	0.00%	0.00%	7.89%	0.002461	Inf
FILIP1	3.26%	1.69%	13.16%	0.003257	8.704111
BIRC6	5.43%	4.22%	18.42%	0.003838	5.078415
DYSF	6.52%	0.84%	10.53%	0.003845	13.58559
DOT1L	1.09%	0.84%	10.53%	0.003845	13.58559
PCDHGC3	0.00%	0.84%	10.53%	0.003845	13.58559
HCN1	0.00%	1.27%	10.53%	0.008073	9.04998
ATRX	1.09%	1.27%	10.53%	0.008073	9.04998
ACIN1	2.17%	1.27%	10.53%	0.008073	9.04998
AP2A2	0.00%	0.42%	7.89%	0.008893	19.83193
ABCC10	2.17%	0.42%	7.89%	0.008893	19.83193
CCDC158	0.00%	0.42%	7.89%	0.008893	19.83193
PTPN23	2.17%	0.42%	7.89%	0.008893	19.83193
ILDR1	0.00%	0.42%	7.89%	0.008893	19.83193
RASGRF2	0.00%	0.42%	7.89%	0.008893	19.83193
GLI2	1.09%	2.53%	13.16%	0.009638	5.77279
ZNF676	0.00%	1.69%	10.53%	0.014535	6.772259
RALGAPA1	1.09%	1.69%	10.53%	0.014535	6.772259
DQX1	2.17%	0.00%	5.26%	0.01866	Inf
ADAM20	0.00%	0.00%	5.26%	0.01866	Inf
LAD1	0.00%	0.00%	5.26%	0.01866	Inf
NPY1R	1.09%	0.00%	5.26%	0.01866	Inf
ZNF391	0.00%	0.00%	5.26%	0.01866	Inf
BTNL9	1.09%	0.00%	5.26%	0.01866	Inf
PTBP1	2.17%	0.00%	5.26%	0.01866	Inf
AATF	0.00%	0.00%	5.26%	0.01866	Inf
ALDH16A1	1.09%	0.00%	5.26%	0.01866	Inf
AQP6	0.00%	0.00%	5.26%	0.01866	Inf
ARAP1	2.17%	0.00%	5.26%	0.01866	Inf
C1QTNF4	0.00%	0.00%	5.26%	0.01866	Inf
CNBD2	0.00%	0.00%	5.26%	0.01866	Inf
FAM184B	2.17%	0.00%	5.26%	0.01866	Inf
FAM98A	2.17%	0.00%	5.26%	0.01866	Inf
FBXO25	0.00%	0.00%	5.26%	0.01866	Inf
GATB	1.09%	0.00%	5.26%	0.01866	Inf
IGJ	0.00%	0.00%	5.26%	0.01866	Inf
KIAA0195	0.00%	0.00%	5.26%	0.01866	Inf
LMNTD2	0.00%	0.00%	5.26%	0.01866	Inf
LRRC72	0.00%	0.00%	5.26%	0.01866	Inf
MIPEP	0.00%	0.00%	5.26%	0.01866	Inf
NR2C2	0.00%	0.00%	5.26%	0.01866	Inf
E2F3	0.00%	0.00%	5.26%	0.01866	Inf

Appendix B: Top 50 genes significantly mutated in Low-HORMAD2 patients. LH, Low HORMAD2; NH, Normal HORMAD2; HH, High HORMAD2; OR, Odd's Ratio.

	Mutation Frequency			Fisher's Exact Testing	
Gene	**LH (n=92)**	**NH (n=237)**	**HH (n=38)**	**p**	**OR (LH vs NH)**
BAP1	14.13%	2.53%	2.63%	1.85E-04	6.29244642
SOCS6	5.43%	0.00%	0.00%	0.001578	Inf
SUPT20HL2	5.43%	0.00%	0.00%	0.001578	Inf
TAF3	5.43%	0.00%	0.00%	0.001578	Inf
TENM4	10.87%	2.11%	2.63%	0.001604	5.622799417
MAP4K5	6.52%	0.42%	0.00%	0.00229	16.32025428
EML6	6.52%	0.42%	0.00%	0.00229	16.32025428
SLC39A8	4.35%	0.00%	2.63%	0.005829	Inf
USP19	4.35%	0.00%	2.63%	0.005829	Inf
MAP7D3	4.35%	0.00%	0.00%	0.005829	Inf
DYSF	6.52%	0.84%	10.53%	0.007076	8.135910749
PRPF39	5.43%	0.42%	2.63%	0.007351	13.44335103
EIF4G1	5.43%	0.42%	0.00%	0.007351	13.44335103
FLNB	6.52%	1.27%	2.63%	0.016421	5.408642382
KMT2E	6.52%	1.27%	5.26%	0.016421	5.408642382
DCHS1	9.78%	2.95%	2.63%	0.018466	3.546532538
CLTCL1	5.43%	0.84%	5.26%	0.020005	6.705639678
DSEL	5.43%	0.84%	0.00%	0.020005	6.705639678
SYNJ1	5.43%	0.84%	2.63%	0.020005	6.705639678
TCEAL6	5.43%	0.84%	0.00%	0.020005	6.705639678
HLA-B	3.26%	0.00%	0.00%	0.021353	Inf
JADE2	3.26%	0.00%	0.00%	0.021353	Inf
ARHGAP24	3.26%	0.00%	2.63%	0.021353	Inf
PPAT	3.26%	0.00%	0.00%	0.021353	Inf
FNDC3B	3.26%	0.00%	2.63%	0.021353	Inf
PCIF1	3.26%	0.00%	0.00%	0.021353	Inf
HK2	3.26%	0.00%	0.00%	0.021353	Inf
PHKA2	3.26%	0.00%	0.00%	0.021353	Inf
IL17RE	3.26%	0.00%	0.00%	0.021353	Inf
CBLB	3.26%	0.00%	0.00%	0.021353	Inf
XYLT1	3.26%	0.00%	0.00%	0.021353	Inf
EPN1	3.26%	0.00%	0.00%	0.021353	Inf
CNIH2	3.26%	0.00%	0.00%	0.021353	Inf
TXN2	3.26%	0.00%	0.00%	0.021353	Inf
DENND6A	3.26%	0.00%	0.00%	0.021353	Inf
MYO1G	3.26%	0.00%	2.63%	0.021353	Inf
DNAJC22	3.26%	0.00%	2.63%	0.021353	Inf
THBD	3.26%	0.00%	0.00%	0.021353	Inf
CS	3.26%	0.00%	2.63%	0.021353	Inf
GRK7	3.26%	0.00%	0.00%	0.021353	Inf
PSRC1	3.26%	0.00%	0.00%	0.021353	Inf
BCAR1	3.26%	0.00%	0.00%	0.021353	Inf
TTLL7	3.26%	0.00%	0.00%	0.021353	Inf
MYLK4	3.26%	0.00%	0.00%	0.021353	Inf
INO80	4.35%	0.42%	5.26%	0.022833	10.64023993
XPO1	4.35%	0.42%	0.00%	0.022833	10.64023993
FLNC	4.35%	0.42%	0.00%	0.022833	10.64023993
MCCC1	4.35%	0.42%	0.00%	0.022833	10.64023993
COBLL1	4.35%	0.42%	2.63%	0.022833	10.64023993
ZNF578	4.35%	0.42%	5.26%	0.022833	10.64023993

Appendix C: Positively and negatively correlated genes with HORMAD2 expression. Included is the positively and negatively-correlated top 50 genes, as well as select genes of interest referenced in the text.

Positive Correlations			Negative Correlations		
	Spearman				
Gene	r	p	Gene	r	p
Top 50			Top 50		
HORMAD2	1	0	TTLL4	-0.53241	3.42E-32
LOC255167	0.682666	4.67E-59	MFSD10	-0.53216	3.70E-32
SEC14L2	0.64746	2.19E-51	LZTS2	-0.50958	3.35E-29
CYP8B1	0.641936	2.84E-50	PLXNA1	-0.50606	9.25E-29
GLYAT	0.638736	1.22E-49	TCF3	-0.50041	4.61E-28
SCP2	0.63161	2.97E-48	PNMA1	-0.50009	5.05E-28
NR1I2	0.63151	3.11E-48	MTHFD1L	-0.49666	1.32E-27
PCK2	0.631084	3.75E-48	CLIC1	-0.49585	1.65E-27
SLC10A1	0.628955	9.55E-48	STX6	-0.49031	7.56E-27
ACADL	0.625482	4.33E-47	LPCAT1	-0.48898	1.09E-26
CYP4F2	0.62197	1.95E-46	DBNDD2	-0.48896	1.09E-26
GBP7	0.619382	5.86E-46	PRMT1	-0.48607	2.38E-26
CYP2A6	0.610313	2.55E-44	C2orf29	-0.47722	2.48E-25
DAO	0.603333	4.27E-43	MARCKS	-0.47574	3.65E-25
GLYATL1	0.602907	5.07E-43	DNMT3A	-0.47472	4.76E-25
ACSM5	0.602177	6.78E-43	BEND3	-0.47405	5.65E-25
PGRMC1	0.597427	4.41E-42	TEAD2	-0.47312	7.19E-25
ETNK2	0.597226	4.77E-42	NAP1L1	-0.4717	1.04E-24
ACSM2A	0.595597	8.99E-42	SRC	-0.47168	1.04E-24
TPPP2	0.594155	1.57E-41	LIMK1	-0.471	1.24E-24
SLC22A1	0.59409	1.61E-41	SLC7A1	-0.47085	1.29E-24
USH2A	0.590814	5.67E-41	EPS8L3	-0.46815	2.56E-24
NFIA	0.590093	7.47E-41	CDCA7	-0.46637	4.00E-24
CYP2A7	0.588622	1.31E-40	CPSF6	-0.4634	8.44E-24
ACSL1	0.58789	1.72E-40	TMEM136	-0.46297	9.40E-24
ECHDC2	0.587657	1.88E-40	RNF24	-0.46292	9.52E-24
HAO2	0.587155	2.27E-40	YEATS2	-0.46008	1.92E-23
RDH16	0.584576	5.99E-40	PHF21A	-0.46005	1.94E-23
GYS2	0.583791	8.03E-40	SMARCA4	-0.45787	3.31E-23
CAT	0.582911	1.11E-39	MPZL1	-0.45626	4.90E-23
MOGAT2	0.582821	1.15E-39	PLEKHG2	-0.45599	5.24E-23
SLC27A5	0.5811	2.18E-39	TYRO3	-0.4557	5.62E-23
FMO4	0.580555	2.67E-39	MTA3	-0.45537	6.09E-23
TAT	0.579366	4.13E-39	MARCKSL1	-0.45416	8.17E-23
UGT2B7	0.578529	5.61E-39	MPV17	-0.45205	1.36E-22
CMBL	0.574423	2.50E-38	GYG1	-0.45187	1.42E-22
TMEM56	0.574005	2.90E-38	DBN1	-0.4518	1.44E-22
ST3GAL6	0.573122	3.99E-38	METTL9	-0.45164	1.50E-22
HSD17B6	0.572453	5.07E-38	NDRG3	-0.45009	2.17E-22
RTP3	0.571436	7.29E-38	MAPRE1	-0.44977	2.35E-22
CYP2C8	0.570527	1.01E-37	MMP11	-0.44911	2.75E-22
METTL7A	0.56879	1.87E-37	GIT1	-0.44897	2.83E-22
GNE	0.568505	2.07E-37	TRIM28	-0.44807	3.51E-22
SLC16A2	0.567828	2.63E-37	CNOT6	-0.44772	3.81E-22
GBA3	0.565956	5.07E-37	IGDCC4	-0.44724	4.27E-22
ACSM2B	0.56582	5.32E-37	TPD52L2	-0.44704	4.48E-22
GADD45A	0.564209	9.34E-37	TESC	-0.44626	5.39E-22
ABCA6	0.563694	1.12E-36	C12orf32	-0.44541	6.57E-22
TTPAL	0.563416	1.23E-36	BZW2	-0.44526	6.82E-22
MPDZ	0.563043	1.40E-36	ARID3A	-0.444	9.16E-22
Genes of Interest			Genes of Interest		
ESR1	0.534461	1.80E-32	TP53	-0.31088	6.93E-11
PPARA	0.475597	3.79E-25			
RXRA	0.309228	8.84E-11			

Appendix D: OPC Smoke vs Control: Top 50 up-regulated genes and enriched pathways.

Top 50 DE Genes				Top 50 DE Pathways		
Gene	**baseMean**	**log2FoldChange**	**padj**	**GO ID**	**Description**	**log10 p-value**
CYP1A1	6958.426	3.829207	0	GO:0030334	regulation of cell migration	-20.8861
IL1RL1	1922.636	3.194871	9.46E-152	GO:0043062	extracellular structure organization	-20.0088
GDA	4159.193	1.470033	2.45E-142	GO:0030198	extracellular matrix organization	-19.7033
TIPARP	4268.382	1.630628	1.89E-132	GO:0032101	regulation of response to external stimulus	-18.6882
TSC22D1	2918.22	1.582869	7.99E-115	GO:0032502	developmental process	-18.1574
CYP1B1	1290.801	2.549338	2.78E-113	GO:0051239	regulation of multicellular organismal process	-17.3143
PTGS2	3872.716	2.740254	2.16E-101	GO:0007165	signal transduction	-17.1972
IER3	1836.272	1.822443	3.59E-95	GO:0007166	cell surface receptor signaling pathway	-17.0123
STC1	1731.408	3.019774	1.82E-91	GO:0040011	locomotion	-16.3925
PHLDA1	4020.224	1.829703	1.51E-90	GO:0042221	response to chemical	-16.1884
DUSP4	1718.406	1.996927	1.81E-89	GO:0070887	cellular response to chemical stimulus	-14.308
PXDN	3181.876	1.678711	7.71E-89	GO:0008284	positive regulation of cell proliferation	-13.6737
CD55	5441.766	1.528854	1.03E-84	GO:0006954	inflammatory response	-13.6003
SERPINE1	1627.351	1.695208	2.33E-79	GO:0048584	positive regulation of response to stimulus	-13.3757
MMP1	1632.595	1.653995	1.76E-72	GO:0048519	negative regulation of biological process	-13.063
AREG	3256.378	1.267105	4.00E-66	GO:0032501	multicellular organismal process	-13.0526
NDRG1	5026.3	1.445248	1.03E-62	GO:0006928	movement of cell or subcellular component	-13.0496
HAS3	2423.441	1.316425	8.63E-62	GO:0048856	anatomical structure development	-12.8729
TXNIP	6457.376	1.277301	7.85E-57	GO:0045765	regulation of angiogenesis	-12.6925
FAM107B	1862.214	1.694014	6.81E-56	GO:0051174	regulation of phosphorus metabolic process	-12.6536
CEMIP2	1176.671	1.643811	7.75E-52	GO:0006952	defense response	-12.4486
NRCAM	508.7299	2.083597	1.20E-50	GO:0003008	system process	-12.4078
LINC00673	1141.612	1.365688	3.59E-50	GO:0022610	biological adhesion	-12.3546
TNFAIP3	5395.723	1.153252	3.19E-49	GO:0019221	cytokine-mediated signaling pathway	-12.032
CD44	17297.05	1.161791	6.13E-47	GO:0048583	regulation of response to stimulus	-11.9547
NOS1	381.6903	2.856612	4.26E-46	GO:0032879	regulation of localization	-11.8794
CREG2	505.662	1.962704	3.24E-45	GO:0048518	positive regulation of biological process	-11.8013
ADAM8	3881.333	1.456852	3.24E-45	GO:0023051	regulation of signaling	-11.7545
ABLIM1	2059.75	1.385381	6.76E-45	GO:0030003	cellular cation homeostasis	-11.5072
SLFN5	4671.576	1.169173	7.77E-45	GO:0050793	regulation of developmental process	-11.4237
OTUB2	733.7474	1.791074	9.06E-44	GO:0009653	anatomical structure morphogenesis	-11.4023
NAV3	907.4259	1.896454	2.92E-43	GO:0010646	regulation of cell communication	-11.3788
SPOCK1	2282.892	1.44425	4.18E-43	GO:0048522	positive regulation of cellular process	-11.289
DUSP1	1421.392	1.529309	1.37E-42	GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	-11.1524
MMP10	788.5641	1.672916	1.76E-41	GO:0050896	response to stimulus	-10.8539
LAMA3	4261.289	1.274821	4.27E-41	GO:0032787	monocarboxylic acid metabolic process	-10.7986
DAPP1	1352.178	1.604376	4.48E-41	GO:0007186	G-protein coupled receptor signaling pathway	-10.6615
EREG	5957.741	1.071411	1.20E-40	GO:0009605	response to external stimulus	-10.5935
SERPINB2	323.6006	3.223971	5.99E-39	GO:0045597	positive regulation of cell differentiation	-10.5467
ITGA2	3012.453	1.403729	9.48E-37	GO:0030155	regulation of cell adhesion	-10.342
FOSL1	9411.858	1.098933	7.29E-34	GO:0050794	regulation of cellular process	-10.2933
CLU	1970.721	1.154701	4.09E-33	GO:0007267	cell-cell signaling	-10.2684
NT5E	2980.893	1.617183	6.73E-33	GO:0048523	negative regulation of cellular process	-10.1135
ADAM12	305.5329	2.15993	2.23E-31	GO:0050789	regulation of biological process	-10.082
ITGA5	5118.159	1.17988	6.29E-31	GO:0065009	regulation of molecular function	-10.0395
IFIT2	5218.308	1.013762	6.61E-31	GO:0070482	response to oxygen levels	-9.8601
FHL1	1844.642	1.314549	6.61E-31	GO:0030001	metal ion transport	-9.6635
OSBP2	2553.224	1.048414	3.67E-30	GO:0001817	regulation of cytokine production	-9.4401
CDKN1A	3770.783	1.172122	7.08E-30	GO:0009607	response to biotic stimulus	-9.3904
ESRG	12070.57	1.139147	3.14E-29	GO:0065008	regulation of biological quality	-9.3449

Appendix E: OPC Smoke vs Control: Top 50 down-regulated genes and enriched pathways.

Top 50 DE Genes				Top 50 DE Pathways		
Gene	**baseMean**	**log2 FoldChange**	**padj**	**GO ID**	**Description**	**log10 p-value**
SCD	9792.015	-1.20041	1.19E-61	GO:0070268	cornification	-12.9355
LBH	678.722	-1.67687	2.21E-58	GO:0000079	regulation of cyclin-dependent protein serine/threonine kinase activity	-12.4572
TNS3	3646.957	-1.09329	2.92E-57	GO:0043087	regulation of GTPase activity	-10.6345
GLDC	327.2592	-1.85882	2.07E-42	GO:0044772	mitotic cell cycle phase transition	-10.3757
INSIG1	1157.118	-1.30832	5.48E-34	GO:0008219	cell death	-10.202
KRT8	75145.8	-1.11751	1.89E-31	GO:0050810	regulation of steroid biosynthetic process	-9.055
PKP1	1576.925	-1.19691	2.01E-29	GO:0051056	regulation of small GTPase mediated signal transduction	-8.7905
NRARP	2524.493	-1.1403	9.57E-29	GO:0006637	acyl-CoA metabolic process	-8.7645
HMGCS1	4468.859	-0.86066	2.69E-27	GO:0032502	developmental process	-8.7055
JADE2	6803.009	-0.84864	2.63E-24	GO:0090383	phagosome acidification	-8.6904
KRT18	45454.19	-0.9719	1.24E-23	GO:0033572	transferrin transport	-8.6757
TNFAIP2	14750.17	-1.07666	1.33E-23	GO:0072512	trivalent inorganic cation transport	-8.6757
FSTL1	1405.875	-0.96787	2.63E-23	GO:0055114	oxidation-reduction process	-8.6234
EPAS1	2447.08	-0.87685	1.59E-21	GO:0006629	lipid metabolic process	-8.4672
PYCR1	2391.665	-0.95117	2.39E-19	GO:0009653	anatomical structure morphogenesis	-8.2291
DHCR7	2837.194	-0.84922	7.57E-19	GO:0006413	translational initiation	-8.2027
KRT5	110598.9	-0.74259	1.51E-18	GO:0044281	small molecule metabolic process	-7.6308
ARPIN	3677.048	-0.7897	3.07E-18	GO:0051493	regulation of cytoskeleton organization	-7.4989
PCSK9	536.0273	-1.3354	5.48E-18	GO:0032501	multicellular organismal process	-7.2366
KRT75	224.0272	-1.93283	7.14E-18	GO:0048856	anatomical structure development	-7.1656
CACNG4	102.2928	-2.0442	3.67E-17	GO:0048732	gland development	-7.1643
KLK6	2912.497	-0.90323	3.79E-17	GO:0009888	tissue development	-7.0535
SAPCD2	1802.69	-0.97352	4.87E-17	GO:0001732	formation of cytoplasmic translation initiation complex	-7.0367
CCNJL	496.7921	-1.23156	5.56E-17	GO:0032787	monocarboxylic acid metabolic process	-6.9318
TGFB2	333.1622	-1.35907	1.27E-16	GO:0019646	aerobic electron transport chain	-6.8697
MTUS1	638.7938	-1.10761	3.83E-16	GO:0006614	SRP-dependent cotranslational protein targeting to membrane	-6.767
COBLL1	701.1392	-1.07363	1.44E-15	GO:0008593	regulation of Notch signaling pathway	-6.7447
NGFR	683.3961	-1.0062	2.11E-15	GO:0043603	cellular amide metabolic process	-6.699
SASH1	638.8379	-1.0327	3.89E-15	GO:0017144	drug metabolic process	-6.6271
SELENOH	650.8708	-1.09989	8.69E-13	GO:1902653	secondary alcohol biosynthetic process	-6.556
KLK9	209.8368	-1.53846	1.11E-12	GO:1902600	hydrogen ion transmembrane transport	-6.3778
FBLN1	2454.536	-0.86067	1.87E-12	GO:0034032	purine nucleoside bisphosphate metabolic process	-6.3188
DAGLA	354.1955	-1.20801	3.50E-12	GO:0033865	nucleoside bisphosphate metabolic process	-6.3188
CYTH3	1696.016	-0.74924	4.65E-12	GO:2000027	regulation of organ morphogenesis	-6.224
PRRG4	933.6244	-0.97756	1.31E-11	GO:0035510	DNA dealkylation	-6.1457
KRT80	7810.255	-0.72158	2.73E-11	GO:0007165	signal transduction	-6.1337
SMO	286.9119	-1.16124	2.81E-11	GO:1901137	carbohydrate derivative biosynthetic process	-6.1051
CFI	222.0421	-1.27572	6.86E-11	GO:0030155	regulation of cell adhesion	-5.9747
CSF1	403.456	-1.13743	1.72E-10	GO:0010628	positive regulation of gene expression	-5.8268
CLDN1	3890.894	-0.68309	3.37E-10	GO:0006490	oligosaccharide-lipid intermediate biosynthetic process	-5.7423
SYNPO	338.0621	-1.10514	6.79E-10	GO:0006959	humoral immune response	-5.5901
FGD3	145.9641	-1.47288	1.85E-09	GO:0097194	execution phase of apoptosis	-5.4377
MVK	808.5615	-0.92399	2.37E-09	GO:0050793	regulation of developmental process	-5.3497
NTN1	630.2286	-0.95886	3.98E-09	GO:0006790	sulfur compound metabolic process	-5.3429
BARX2	529.0542	-0.8971	4.22E-09	GO:0045596	negative regulation of cell differentiation	-5.2733
STARD4	803.6638	-0.86791	1.30E-08	GO:0072521	purine-containing compound metabolic process	-5.2565
LOXL4	429.7935	-1.15421	2.27E-08	GO:0008202	steroid metabolic process	-5.2336
SYT12	2739.682	-0.72358	3.61E-08	GO:0006935	chemotaxis	-5.2204
CEBPB	987.8412	-0.90909	5.35E-08	GO:1902903	regulation of supramolecular fiber organization	-5.1925
H2BC9	220.8883	-1.24487	5.70E-08	GO:1902410	mitotic cytokinetic process	-5.1878

Appendix F: OPC Radiation vs Control: Top 50 up-regulated genes and enriched pathways.

Top 50 DE Genes				Top 50 DE Pathways		
Gene	**baseMean**	**log2FoldChange**	**padj**	**GO ID**	**Description**	**log10 p-value**
PTGS2	3872.71623	5.810089466	0	GO:0007165	signal transduction	-40.0052
TNFAIP3	5395.7228	2.827985763	0	GO:0006952	defense response	-33.9914
PHLDA1	4020.22396	3.544802493	0	GO:0007166	cell surface receptor signaling pathway	-31.6819
SLFN5	4671.57637	2.835980231	0	GO:0019221	cytokine-mediated signaling pathway	-29.9586
IFIT2	5218.30785	2.749169448	0	GO:0002376	immune system process	-29.301
ISG20	1593.90368	4.098077799	0	GO:0006950	response to stress	-28.821
TXNIP	6457.37624	2.798820757	0	GO:0048583	regulation of response to stimulus	-26.762
PXDN	3181.87627	3.105260927	0	GO:0043207	response to external biotic stimulus	-26.6596
CDKN1A	3770.78321	3.172287522	0	GO:0051239	regulation of multicellular organismal process	-26.2757
NT5E	2980.89281	4.482130183	0	GO:0009607	response to biotic stimulus	-26.1618
ITGA5	5118.15916	3.045620557	0	GO:0009605	response to external stimulus	-24.9172
STC1	1731.40823	5.197824624	0	GO:0006955	immune response	-24.7905
CD55	5441.76636	2.557932712	2.12E-304	GO:0050896	response to stimulus	-24.5317
KLF6	2439.72378	2.727239277	1.04E-300	GO:0030334	regulation of cell migration	-23.4698
IER3	1836.27197	2.835794299	1.60E-292	GO:0048519	negative regulation of biological process	-22.5638
DUSP1	1421.392	3.267481865	5.83E-291	GO:0048584	positive regulation of response to stimulus	-22.1331
IL1RL1	1922.63554	4.170260623	4.43E-283	GO:0042221	response to chemical	-20.5867
ADAM8	3881.33326	3.069772703	1.21E-282	GO:0043062	extracellular structure organization	-20.2798
LCP1	3078.36145	3.319732306	3.54E-274	GO:0003008	system process	-19.7447
FN1	3343.53914	2.916896628	5.35E-258	GO:0030198	extracellular matrix organization	-19.7122
HLA-B	9279.29887	2.236540555	1.42E-248	GO:0050789	regulation of biological process	-19.3335
NDRG1	5026.3003	2.461031425	5.75E-241	GO:0050794	regulation of cellular process	-19.3188
LIF	2319.22273	2.647623737	4.42E-238	GO:0070887	cellular response to chemical stimulus	-19.2676
HBEGF	1720.80538	3.727950102	2.73E-235	GO:0010646	regulation of cell communication	-19.015
CXCL8	1148.14842	3.308487999	3.35E-233	GO:0048518	positive regulation of biological process	-18.7447
ATF3	1232.15588	3.328525785	5.87E-227	GO:0023051	regulation of signaling	-18.6383
SH3KBP1	2718.92369	2.302166241	6.61E-218	GO:0032501	multicellular organismal process	-18.2668
IFIT3	4545.91645	2.155645512	3.70E-217	GO:0022610	biological adhesion	-18.172
DDX58	3631.36607	2.357709203	3.52E-215	GO:0006954	inflammatory response	-18.1152
SPOCK1	2282.89233	2.712760713	2.44E-212	GO:0032502	developmental process	-18.0372
MAP1B	1685.86264	2.786824564	4.05E-205	GO:0023052	signaling	-17.6459
PLAU	16238.4062	1.824138502	2.08E-193	GO:0032879	regulation of localization	-17.5452
NCF2	773.657154	5.162400294	3.67E-193	GO:0065007	biological regulation	-17.5171
HDAC9	667.630514	3.720057215	1.28E-179	GO:0048523	negative regulation of cellular process	-17.0195
OASL	7551.37186	2.073487487	7.08E-178	GO:0065009	regulation of molecular function	-16.8477
ITGA2	3012.45337	2.635717626	1.70E-177	GO:0034097	response to cytokine	-16.7825
PPP1R15A	1579.25516	2.44170598	9.34E-177	GO:0048522	positive regulation of cellular process	-16.5272
STC2	4096.65485	1.54885255	2.22E-175	GO:0040011	locomotion	-16.2487
LAMC2	24155.4788	2.255290374	3.43E-174	GO:0030155	regulation of cell adhesion	-16.0259
VEGFA	4743.94255	1.68943676	4.46E-173	GO:0007186	G-protein coupled receptor signaling pathway	-15.2604
LCN2	1017.01631	3.600987257	2.81E-172	GO:0050793	regulation of developmental process	-15.0804
TNC	2177.30292	2.249336183	4.00E-169	GO:0007154	cell communication	-14.821
AHNAK2	1020.27789	2.638322321	1.10E-167	GO:0007267	cell-cell signaling	-14.6904
SDC4	11643.8748	1.616254732	3.95E-167	GO:0001817	regulation of cytokine production	-14.5654
DAPP1	1352.17813	2.765486986	1.42E-162	GO:0043901	negative regulation of multi-organism process	-14.5638
DUSP4	1718.40594	2.521525994	1.17E-161	GO:0043900	regulation of multi-organism process	-14.2182
ICAM1	2150.79427	2.192215751	1.52E-161	GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	-14.0155
IL6	576.586089	3.884871197	1.46E-158	GO:0051174	regulation of phosphorus metabolic process	-13.8894
PPP1R18	2994.18555	1.626781237	3.62E-156	GO:0048856	anatomical structure development	-13.5214
SOD2	12570.239	1.819557108	4.44E-151	GO:0051246	regulation of protein metabolic process	-13.3054
Additional Genes of Interest						
CD44	17297.048	1.627720621	5.69E-116			

Appendix G: OPC Radiation vs Control: Top 50 down-regulated genes and enriched pathways.

Top 50 DE Genes				Top 50 DE Pathways		
Gene	**baseMean**	**log2FoldChange**	**padj**	**GO ID**	**Description**	**log10 p-value**
KRT5	110598.9262	-2.150624826	0	GO:1901360	organic cyclic compound metabolic process	-51.0367
GPX2	1156.019819	-3.592025108	0	GO:0046483	heterocycle metabolic process	-50.4179
TNS3	3646.956969	-2.107452922	1.35E-296	GO:0006725	cellular aromatic compound metabolic process	-48.5186
KLK6	2912.496783	-2.743884543	1.23E-278	GO:0006139	nucleobase-containing compound metabolic process	-48.4685
KRT15	6198.24148	-2.149599034	4.84E-223	GO:0006614	SRP-dependent cotranslational protein targeting to membrane	-47.384
KRT13	14899.16923	-3.981729255	7.87E-203	GO:0006413	translational initiation	-45.4828
FAT2	3862.65102	-2.578028283	4.63E-188	GO:0008152	metabolic process	-43.3726
SYT12	2739.681866	-2.266035951	4.24E-186	GO:0019083	viral transcription	-42.2343
S100A10	9420.734092	-1.540013305	5.26E-178	GO:0044237	cellular metabolic process	-41.4976
NRARP	2524.492915	-2.274148382	4.78E-160	GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	-38.4672
ZNF488	961.5806036	-2.711968546	1.02E-155	GO:0071704	organic substance metabolic process	-34.6108
PKP1	1576.925237	-2.342414883	7.72E-151	GO:0009058	biosynthetic process	-32.5229
UGT1A6	1054.379934	-3.030155281	8.01E-141	GO:0044238	primary metabolic process	-32.2798
KIF20A	3439.864876	-1.564917792	2.10E-139	GO:1901576	organic substance biosynthetic process	-32.2161
KLK10	1793.571667	-2.575960775	4.28E-132	GO:0009987	cellular process	-30.6038
SCD	9792.014528	-1.546958218	1.10E-121	GO:0016071	mRNA metabolic process	-28.4609
ALDH3A1	4038.0789	-3.685367671	1.58E-116	GO:0043603	cellular amide metabolic process	-26.8996
TNNI2	793.1610459	-3.237699134	8.03E-113	GO:0043933	macromolecular complex subunit organization	-26.6799
KRT4	3244.801801	-7.366226162	4.27E-111	GO:0051301	cell division	-25.4802
CAV1	33015.52095	-1.438549726	6.18E-108	GO:0006807	nitrogen compound metabolic process	-25.2449
TRIM16L	2847.46316	-1.550376906	4.15E-105	GO:0046700	heterocycle catabolic process	-24.6326
RARG	3769.949271	-1.470340434	5.03E-105	GO:0051276	chromosome organization	-23.8356
FGFR2	996.5579288	-1.960237313	2.28E-102	GO:0022904	respiratory electron transport chain	-22.857
SLC47A2	435.7677092	-3.811786716	2.53E-99	GO:0010564	regulation of cell cycle process	-21.7328
PAX9	2093.386099	-1.549989047	3.83E-99	GO:0007059	chromosome segregation	-20.8962
PTMA	22341.62713	-1.162283045	2.03E-97	GO:0044281	small molecule metabolic process	-20.8539
LMNB1	5913.227587	-1.297295991	6.18E-96	GO:0055114	oxidation-reduction process	-20.0773
KLK5	4209.861564	-1.771063719	4.84E-94	GO:0032981	mitochondrial respiratory chain complex I assembly	-19.3936
HMGCS1	4468.859123	-1.295306958	7.87E-91	GO:0071840	cellular component organization or biogenesis	-18.5436
INSIG1	1157.117821	-1.924482758	1.43E-89	GO:0006259	DNA metabolic process	-17.8356
CAV2	11013.27349	-1.416884137	1.86E-88	GO:0016043	cellular component organization	-17.279
GLDC	327.2591536	-2.563955418	1.68E-86	GO:0043170	macromolecule metabolic process	-16.7905
TRIM16	5389.860757	-1.250606941	2.94E-78	GO:0051186	cofactor metabolic process	-16.3757
SERPINB13	460.5407964	-2.873120677	1.05E-76	GO:0009117	nucleotide metabolic process	-15.1726
HMGB1	24698.99691	-1.216697062	8.17E-76	GO:0006091	generation of precursor metabolites and energy	-14.4461
G6PD	7252.432966	-1.300131248	1.61E-73	GO:0044260	cellular macromolecule metabolic process	-13.618
JADE2	6803.008813	-1.203043687	1.08E-69	GO:0071826	ribonucleoprotein complex subunit organization	-13.3675
RNPEP	3115.876311	-1.246318039	2.23E-69	GO:0006396	RNA processing	-13.3072
ANXA8L1	4353.274464	-1.18409185	3.28E-69	GO:0072521	purine-containing compound metabolic process	-13.2069
HMGN2	17286.58068	-0.908827086	5.86E-68	GO:0010608	posttranscriptional regulation of gene expression	-12.9245
KLK9	209.8368098	-3.499893864	6.80E-68	GO:0006732	coenzyme metabolic process	-12.7852
TKT	17286.6192	-1.227083282	2.97E-67	GO:0006281	DNA repair	-12.4237
CDCA7	3817.347564	-1.214795622	1.95E-66	GO:0051704	multi-organism process	-12.3279
ANXA8	9162.403211	-1.073358314	3.17E-66	GO:0006333	chromatin assembly or disassembly	-12.0467
GCLC	3820.507442	-1.212807054	5.53E-66	GO:0017144	drug metabolic process	-11.8386
SAPCD2	1802.69024	-1.581226249	1.50E-64	GO:0051641	cellular localization	-11.0773
DSC3	5574.175705	-1.22492553	1.73E-64	GO:0009056	catabolic process	-10.8477
LOC102724852	22799.38875	-3.921260718	2.23E-63	GO:0007049	cell cycle	-10.5834
PDLIM1	4007.681069	-1.235004963	8.08E-63	GO:0043488	regulation of mRNA stability	-10.433
ALDH3A2	4808.08697	-1.135099958	3.91E-62	GO:0006325	chromatin organization	-10.4237

Appendix H: OPC Radiation and Smoke vs Control: Top 50 up-regulated genes and enriched pathways.

Top 50 DE Genes				Top 50 DE Pathways		
Gene	**baseMean**	**log2FoldChange**	**padj**	**GO ID**	**Description**	**log10 p-value**
PTGS2	3872.716	6.297083	0	GO:0007165	signal transduction	-34.7011
NDRG1	5026.3	3.60161	0	GO:0006952	defense response	-29.8069
PHLDA1	4020.224	3.816223	0	GO:0019221	cytokine-mediated signaling pathway	-27.8794
CDKN1A	3770.783	3.651169	0	GO:0007166	cell surface receptor signaling pathway	-27.1024
ITGA5	5118.159	3.485716	0	GO:0002376	immune system process	-25.6162
SLFN5	4671.576	2.888488	0	GO:0006950	response to stress	-23.6556
TXNIP	6457.376	2.907708	0	GO:0009607	response to biotic stimulus	-23.4841
ISG20	1593.904	4.012713	0	GO:0043207	response to external biotic stimulus	-23.2993
IER3	1836.272	3.207619	0	GO:0070887	cellular response to chemical stimulus	-23.1518
PLAU	16238.41	2.365965	0	GO:0051239	regulation of multicellular organismal process	-22.4295
PXDN	3181.876	2.954099	0	GO:0050896	response to stimulus	-22.0675
TNFAIP3	5395.723	2.473732	0	GO:0042221	response to chemical	-21.9245
STC2	4096.655	2.049309	0	GO:0048583	regulation of response to stimulus	-21.6091
CD55	5441.766	2.698811	0	GO:0048519	negative regulation of biological process	-21.5317
IL1RL1	1922.636	4.523533	0	GO:0030334	regulation of cell migration	-20.684
LCP1	3078.361	3.612691	0	GO:0048584	positive regulation of response to stimulus	-20.4789
STC1	1731.408	5.201162	0	GO:0023052	signaling	-20.3605
SH3KBP1	2718.924	2.692268	0	GO:0009605	response to external stimulus	-20.1824
HBEGF	1720.805	4.224866	0	GO:0022610	biological adhesion	-19.3778
NT5E	2980.893	4.114553	0	GO:0050789	regulation of biological process	-18.8013
ADAM8	3881.333	3.193388	0	GO:0003008	system process	-18.7545
IFIT2	5218.308	2.41918	0	GO:0032501	multicellular organismal process	-18.7282
SPOCK1	2282.892	3.152767	8.97E-299	GO:0040011	locomotion	-18.1733
DUSP1	1421.392	3.306959	5.95E-297	GO:0048518	positive regulation of biological process	-18.0799
KLF6	2439.724	2.714543	1.84E-295	GO:0023051	regulation of signaling	-17.8794
HLA-B	9279.299	2.313757	1.18E-268	GO:0010646	regulation of cell communication	-17.7328
ATF3	1232.156	3.569811	1.45E-264	GO:0050794	regulation of cellular process	-17.6216
VEGFA	4743.943	1.981302	8.68E-255	GO:0048523	negative regulation of cellular process	-17.52
LCN2	1017.016	4.289111	4.63E-254	GO:0007267	cell-cell signaling	-17.4522
TIPARP	4268.382	2.100149	8.25E-252	GO:0065009	regulation of molecular function	-16.618
MAP1B	1685.863	2.979005	7.93E-238	GO:0043062	extracellular structure organization	-16.5143
LAMB3	22810.73	1.981575	1.11E-228	GO:0006954	inflammatory response	-16.3354
PPP1R18	2994.186	1.878806	7.20E-222	GO:0030155	regulation of cell adhesion	-16.2132
DUSP4	1718.406	2.870988	1.87E-217	GO:0030198	extracellular matrix organization	-16.0487
CYP1A1	6958.426	2.958687	2.39E-216	GO:0032879	regulation of localization	-15.9666
FN1	3343.539	2.705632	2.95E-216	GO:0007154	cell communication	-15.8894
LAMC2	24155.48	2.467795	1.36E-214	GO:0034097	response to cytokine	-15.6383
AHNAK2	1020.278	2.916117	2.56E-210	GO:0048522	positive regulation of cellular process	-15.58
DAPP1	1352.178	3.09403	1.31E-209	GO:0006811	ion transport	-15.3585
UPP1	2123.356	2.444543	1.16E-208	GO:0007186	G-protein coupled receptor signaling pathway	-14.7496
NCF2	773.6572	5.325266	3.30E-206	GO:0065007	biological regulation	-14.5302
TSC22D1	2918.22	1.959555	2.51E-200	GO:0001817	regulation of cytokine production	-14.015
PPP1R15A	1579.255	2.575283	6.59E-199	GO:0030001	metal ion transport	-13.8297
FHL1	1844.642	2.735852	1.60E-198	GO:0051174	regulation of phosphorus metabolic process	-13.7011
LIF	2319.223	2.439275	3.30E-195	GO:1901342	regulation of vasculature development	-13.6459
HDAC9	667.6305	3.866518	6.79E-195	GO:0051246	regulation of protein metabolic process	-13.4168
SDC4	11643.87	1.707889	2.87E-191	GO:0050793	regulation of developmental process	-13.3179
CD24	857.7294	3.214401	8.42E-189	GO:0050731	positive regulation of peptidyl-tyrosine phosphorylation	-13.2716
CEMIP2	1176.671	2.731178	1.25E-186	GO:0098771	inorganic ion homeostasis	-13.2396
VAT1	4530.535	1.762125	4.89E-186	GO:0043086	negative regulation of catalytic activity	-13.2076

Appendix I: OPC Radiation and Smoke vs Control: Top 50 down-regulated genes and enriched pathways.

Top 50 DE Genes				Top 50 DE Pathways		
Gene	**baseMean**	**log2 FoldChange**	**padj**	**GO ID**	**Description**	**log10 p-value**
KRT5	110598.9	-2.36378	0	GO:0006139	nucleobase-containing compound metabolic process	-69.0899
TNS3	3646.957	-2.55738	0	GO:0046483	heterocycle metabolic process	-67.6968
KLK6	2912.497	-2.4513	5.15E-216	GO:1901360	organic cyclic compound metabolic process	-64.2865
KRT15	6198.241	-2.09803	3.82E-208	GO:0006725	cellular aromatic compound metabolic process	-64.2321
SYT12	2739.682	-2.3957	2.19E-204	GO:0044237	cellular metabolic process	-39.9281
KIF20A	3439.865	-1.69535	2.39E-163	GO:0019083	viral transcription	-39.0353
SCD	9792.015	-1.73563	5.55E-160	GO:0008152	metabolic process	-38.5406
PKP1	1576.925	-2.42455	3.91E-157	GO:0051301	cell division	-37.5317
GPX2	1156.02	-2.33353	1.62E-150	GO:0051276	chromosome organization	-37.342
NRARP	2524.493	-2.21736	1.67E-148	GO:0071704	organic substance metabolic process	-35.5258
KRT13	14899.17	-3.43007	6.91E-147	GO:0044238	primary metabolic process	-34.6946
LMNB1	5913.228	-1.5017	2.09E-136	GO:0034645	cellular macromolecule biosynthetic process	-34.6536
IGFL1	1323.214	-2.64062	3.49E-135	GO:0006412	translation	-34.5686
FAT2	3862.651	-2.18879	9.46E-130	GO:0006614	SRP-dependent cotranslational protein targeting to membrane	-34.0506
S100A10	9420.734	-1.34548	1.61E-125	GO:0006807	nitrogen compound metabolic process	-33.083
PTMA	22341.63	-1.26912	6.91E-123	GO:0016071	mRNA metabolic process	-31.1267
ZNF488	961.5806	-2.4291	2.24E-121	GO:0006259	DNA metabolic process	-30.0343
CAV1	33015.52	-1.49163	4.11E-118	GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay	-28.6676
TNNI2	793.161	-3.35808	2.49E-115	GO:0009058	biosynthetic process	-28.056
CDCA7	3817.348	-1.50812	1.77E-112	GO:0043170	macromolecule metabolic process	-27.6364
TMPO	10133.53	-1.23832	7.14E-112	GO:0006281	DNA repair	-25.3645
GLDC	327.2592	-2.89489	5.33E-98	GO:0009987	cellular process	-25.2857
KLK10	1793.572	-2.25144	5.52E-97	GO:0006396	RNA processing	-24.9788
DEPDC1	4119.837	-1.46637	4.99E-96	GO:0007059	chromosome segregation	-24.719
HMGB1	24699	-1.32427	1.19E-94	GO:0010564	regulation of cell cycle process	-23.6364
PAX9	2093.386	-1.52353	1.55E-92	GO:0043933	macromolecular complex subunit organization	-22.0123
SAPCD2	1802.69	-1.85123	3.64E-91	GO:0071840	cellular component organization or biogenesis	-19.9355
HMGB2	5281.305	-1.30573	9.39E-90	GO:0046700	heterocycle catabolic process	-19.8477
KRT4	3244.802	-6.53652	1.54E-89	GO:0044260	cellular macromolecule metabolic process	-19.7595
FGFR2	996.5579	-1.87728	5.57E-89	GO:0043603	cellular amide metabolic process	-19.7328
PRC1	7107.805	-1.26539	6.43E-88	GO:0051983	regulation of chromosome segregation	-18.0026
BIRC5	6689.06	-1.21365	7.14E-87	GO:0071103	DNA conformation change	-17.6925
ARPIN	3677.048	-1.3246	9.54E-84	GO:0016043	cellular component organization	-17.382
KNL1	3145.606	-1.43344	1.03E-82	GO:0006325	chromatin organization	-17.066
HMGCS1	4468.859	-1.24604	8.05E-81	GO:0006996	organelle organization	-16.1599
KREMEN1	1191.185	-1.63623	4.49E-80	GO:0006364	rRNA processing	-15.8153
BRI3BP	3807.455	-1.25934	2.38E-79	GO:0007049	cell cycle	-15.3261
GTSE1	8609.512	-1.04272	5.31E-77	GO:0071826	ribonucleoprotein complex subunit organization	-14.5376
SRSF2	10580.98	-0.95234	3.70E-73	GO:0031055	chromatin remodeling at centromere	-14.057
CENPF	9739.233	-1.47545	4.19E-69	GO:0007010	cytoskeleton organization	-13.3883
MKI67	16806.28	-1.39385	4.36E-69	GO:0034660	ncRNA metabolic process	-13.3565
LBH	678.722	-1.76214	5.31E-68	GO:0007017	microtubule-based process	-13.2612
CCNA2	3678.823	-1.23791	1.36E-67	GO:0006913	nucleocytoplasmic transport	-12.8069
LOXL4	429.7935	-2.83339	1.01E-66	GO:0071824	protein-DNA complex subunit organization	-12.7773
ASPM	6701.882	-1.34648	3.67E-65	GO:0009117	nucleotide metabolic process	-12.5406
DLGAP5	3770.648	-1.2791	3.61E-64	GO:0051704	multi-organism process	-12.3747
MCM6	5610.37	-1.15403	8.32E-64	GO:0010608	posttranscriptional regulation of gene expression	-12.1385
PGAM5	3197.811	-1.14549	2.51E-63	GO:0044281	small molecule metabolic process	-11.9872
PRR11	5586.412	-1.20897	4.73E-63	GO:0033554	cellular response to stress	-11.9788
CAV2	11013.27	-1.24053	2.81E-62	GO:0051641	cellular localization	-11.7799

Appendix J: Expression of various genes of interest which show synergistic changes in gene expression in response to smoke and radiation treatment. Gene expression values are normalised expression values from DESeq2.

	Control			Smoke			Radiation			Smoke and Radiation		
	FaDu1	FaDu2	FaDu3	FaDu4	FaDu5	FaDu6	FaDu7	FaDu8	FaDu9	FaDu10	FaDu11	FaDu12
PTGS2/COX-2	122.2811	85.16342	117.1454	802.647	652.8297	728.7884	6560.863	5811.143	5930.152	9062.994	8394.284	8204.304
MMP9	12.41718	8.291713	5.646436	44.85704	35.97064	31.00327	236.6094	242.8486	230.5339	628.4752	524.0239	500.109
MMP2	76.89681	54.80768	49.79922	82.66552	124.7035	109.3038	184.5488	220.3175	179.2774	260.3474	258.7747	253.0172
ITGA5	906.2721	932.0114	864.4514	2290.64	1822.371	2019.527	7793.96	7176.709	7341.08	10318.29	10382.39	9570.209
VIM	8.271406	15.41107	3.374679	18.49766	52.79317	12.80813	58.01921	43.99112	49.67796	70.24888	69.89065	56.36234
VEGFA	1956.402	1898.822	1985.916	3200.467	2854.302	3123.364	6358.41	6143.151	6339.565	7571.136	7674.875	7820.901
VEGFB	267.0637	220.5582	245.3141	240.6852	249.9348	205.383	205.6559	223.4199	281.0644	314.8785	341.7065	348.7868
VEGFC	252.455	254.7012	237.0254	622.8445	469.8201	592.0603	1371.944	1164.552	1303.582	1522.814	1454.661	1413.119
CLDN4	1477.852	1534.66	1241.469	1073.731	1429.33	1094.335	1919.527	2156.638	2093.591	3124.726	3524.011	3589.609
SLPI	102.0853	152.3243	111.4343	82.40548	135.4689	81.84905	198.6815	174.1682	173.2775	446.4621	509.4567	484.8566
CTSD	11186.02	10456.82	8776.143	7136.675	9014.902	7072.337	8436.032	9000.315	9049.517	12905.08	14261.36	14133.87
CTSA	2964.934	3044.801	2756.306	3328.249	3301.929	3202.216	3976.004	3761.297	3902.454	5122.093	5300.16	5273.353
CD44	6830.924	7176.04	7324.107	17035.93	14262.31	16423.51	22814.02	21396.44	21695.63	24940.67	23654.26	24010.73
RHOB	342.5214	324.5481	303.8395	360.1483	402.4367	379.2699	732.0433	748.2898	813.1401	907.6054	851.0105	900.9422
RHOC	4128.139	4968.782	4404.788	4619.463	5026.69	4571.405	6884.821	6815.814	6900.13	7882.062	8666.781	7391.476
α7nAChR	66.17002	50.09258	53.5058	84.9625	68.3539	78.10667	121.8741	114.0796	124.3403	99.55206	121.6583	87.43975
KRT5	221008.1	226541.7	210405.2	127149.3	136915	129170.4	50504.61	49313.09	48359.26	40182.91	43556.87	44080.7
TNS3	7971.281	7812.417	7619.99	3656.29	3513.515	3794.863	1880.086	1773.266	1772.743	1281.479	1365.465	1322.089

Appendix K: Expression of nicotin acetylcholine receptors in OPC cell lines in response to smoke and radiation treatment