Functional characterization of germline *SMG1* mutations associated with pancreatic ductal adenocarcinoma.

Tatiana Lenko

Department of Human Genetics, McGill University, Montreal

August 2021

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Science

Abstract

Pancreatic ductal adenocarcinoma (PDAC) is the third leading cause of cancer death in Canada. Since genetic and familial risk is thought to underlie 10% of PDAC, surveillance for early cancer detection in at-risk individuals may reduce cancer death rates. However, gene-based surveillance strategies are limited as known PDAC predisposition genes account for only a minority of familial PDAC (FPC). Therefore, there is an unmet clinical need to delineate the hereditary basis for the fraction of FPC unexplained by known predisposition genes. A previous case-control regional gene association study published from our lab identified *SMG1* as a novel predisposition gene with missense variants driving the association with PDAC. I hypothesized that *SMG1* acts as a tumour suppressor gene in PDAC and the missense variants identified in our previous study confer predisposition to PDAC through reduced SMG1 function. The aim of my dissertation was to generate cell lines harbouring the missense mutations identified by Wong et al, and to characterize the resultant cell lines for hallmarks of tumour suppressor and SMG1-specific functions. HEK293T cell lines bearing the c.10921A>G *SMG1* missense mutation were generated using CRISPR gene editing. *SMG1* c.10921A>G cell lines displayed increased migration, decreased proliferation, decreased nonsense-mediated decay, and decreased survival after exposure to ionizing radiation. Our findings suggest that *SMG1* c.10921A>G variant harbours altered functional activity, which may contribute to PDAC predisposition. *In vivo* validation of c.10921A>G *SMG1* functional activity and characterization of additional *SMG1*-PDAC associated variants is needed to strengthen the evidence for *SMG1* as a PDAC predisposition gene.

Résumé

L'adénocarcinome canalaire pancréatique (ADP) est la troisième plus importante cause de décès associée au cancer au Canada. Considérant que le risque génétique et familial est à l'origine d'environ 10% des ADP, une surveillance permettant la détection précoce du cancer chez les personnes à risque pourrait réduire les taux de mortalité causés par cette maladie. Toutefois, les stratégies de surveillance instrumentalisant des gènes de susceptibilité sont encore imparfaites étant donné notre connaissance très limitée des gènes impliqués dans la prédisposition au développement de l'ADP familial. Par conséquent, il apparaît nécessaire de mieux définir les causes héréditaires pouvant expliquer la présence d'ADP familiaux ne pouvant pas être associés à des gènes de prédisposition déjà identifiés. Une précédente étude d'association génétique régionale cas-contrôle publiée par notre laboratoire a montré que des variants faux-sens du gène *SMG1* agissent comme nouveaux facteurs de prédisposition à l'ADP familial. J'ai émis l'hypothèse que *SMG1* agit comme un suppresseur de tumeur dans l'ADP et que la réduction de son activité, causée par les mutations faux-sens identifiées dans notre étude précédente, confère une prédisposition à l'ADP. L'objectif de ce projet était de générer des lignées cellulaires portant les mutations faux-sens décrites par Wong et al. et de caractériser ces lignées cellulaires afin de disséquer les fonctions de *SMG1*. Des lignées cellulaires HEK293T portant la mutation faux-sens c.10921A>G *SMG1* ont été générées en utilisant l'édition génétique par CRISPR. Les lignées cellulaires *SMG1* c.10921A>G ont présenté une augmentation de la migration, une diminution de la prolifération, une diminution de la dégradation des ARNm non-sens ainsi qu'une réduction de la survie suivant l'exposition aux rayonnements ionisants (IR). Nos résultats suggèrent que le variant *SMG1* c.10921A>G présente une activité fonctionnelle altérée qui pourrait contribuer à la prédisposition à l'ADP. La validation *in vivo* de l'activité fonctionnelle de *SMG1* c.10921A>G et la caractérisation d'autres variants de SMG1 associés à l'ADP sont nécessaires afin de mieux décrire et cimenter notre compréhension du rôle de *SMG1* dans la prédisposition à l'ADP.

Table of Contents

Acknowledgements

I would like to first express my appreciation and gratitude to my supervisor Dr. George Zogopoulos. The guidance, knowledge, and opportunities he has given me over my three-year tenure in his lab have invaluable to my growth as a scientist and furthered my interest in medicine.

I want to extend my sincere gratitude to the members of my committee, Dr. Logan Walsh, Dr. Sidong Huang, and Dr. Jerry Pelletier for their advisement and constructive criticism on my project. I would like to especially thank Dr. Jerry Pelletier who went above and beyond in his support as both a committee member and a collaborator on this project.

I feel very fortunate to have overlapped with such kind, intelligent, and motivated graduate students Bryn Golesworthy, Yifan Wang, and Joan Miguel Romero. Their support and friendship throughout my time here has been invaluable; Bryn, Yifan and Joan have all helped me with conceptualizing analysis or assisting with coding in R.

I am grateful to have had honorary lab mates in the Pelletier lab: Samer Girgas, Minza Haque, Leo Shen, Lauren Pugsley, and Sai Naineni, who all provided experimental advice and support. Further, I would like to acknowledge Dr. Francis Roberts of the Pelletier who taught me most of the experimental techniques I learned with patience and enthusiasm.

In addition, I would like to thank the all the members of the Zogopoulos lab over the last three years, especially Céline Domecq and Guillaume Bourdel. As well as Cavin Wong who laid the ground work for this thesis.

I am grateful for the support I received prior to attending McGill, from my professors and peers in the genetics program at University of Western Ontario. I would like to especially acknowledge my honours thesis supervisor Dr. Kathleen Hill, whose encouragement and contagious passion for science led me to apply for graduate studies.

I would like to extend my appreciation to fellow graduate students at the GCRC, RI-MUHC, and in the department of human genetics for the support and friendship over the last three years.

Lastly, I would like to thank my family for their encouragement and support in everything I do, and especially during my graduate work.

Dr. George Zogopoulos provided guidance and expertise on experimental design, data interpretation, and the preparation of this thesis.

Dr. François Fabi assisted with the English-to-French translation of the abstract

Dr. Jerry Pelletier provided guidance in experimental design, technical support, and data analysis.

Mr. Samer Girgis performed PCR to confirm zygosity of mutant c.10291A>G clones.

Dr. Logan Walsh and **Dr. Sidong Huang** provided expertise in CRISPR/Cas9 experimental design.

I performed all experiments presented in this dissertation, unless mentioned above. I prepared this dissertation.

List of Tables and Figures

LIST OF FIGURES

LIST OF TABLES

List of Abbreviations

Abbreviation	Definition
AML	Acute Myeloid Leukemia
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRISPRa	CRISPR activating
CRISPRi	CRISPR inhibiting
dCas9	Dead Cas9
DECID	Decay Inducing Complex
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
DT	Doubling Time
EJC	Exon Junction Complex
EMT	Epithelial-Mesenchymal Transition
FACS	Fluorescence-Activated Cell Sorting
FAT	FRAP ATM and TRAP
FATC	FRAP ATM and TRAP C-Terminal
FBS	Fetal Bovine Serum
FDR	First Degree Relative
FPC	Familial Pancreatic Cancer
GFP	Green Fluorescent Protein
HDR	Homology Directed Repair
HEAT	Huntingtin Elongation factor 3, Alpha-regulatory subunit of protein phosphatase 2A and TOR1
HEK293t	Transformed Human Embryonic Kidney
HNSCC	Head and neck squamous cell carcinoma
IR	Ionizing Radiation
LOO	Leave One Out
miRNA	microRNA
MiST	Mixed Effects Score Test
MMR	Mismatch Repair
mRNA	messenger RNA
NaAs	Sodium Ascorbite
NHEJ	Non-homologous End Joining
NMD	Nonsense-mediated Decay
PCR	Polymerase Chain Reaction
PDAC	Pancreatic Ductal Adenocarcinoma
PIKK	PI3K-related Kinase
PJ	Peutz-Jeghers Syndrome
PTC	Premature Termination Codon
RLU	Relative Light Unit
RNA	Ribonucleic acid
shRNA	short hairpin RNA
siRNA	small interfering RNA
snoRNA	small nucleolar RNA
SMG1c	SMG1 complex
SURF	SMG1-UPF1-eRF1-eRF3

TALEN	Transcription Activator-like Effector Nucleases
TSG	Tumour Suppressor Gene
UV	Ultra Violet
VUS	Variants of unknown significance
WT	Wild Type
ZFN	Zinc Finger Nucleases

Chapter I: Introduction

1.1 Pancreatic Cancer

1.1.1 The Pancreas

The pancreas is a retroperitoneal organ that functions in glucose regulation and digestion. Anatomically, the pancreas is divided into the head, uncinate process, neck, body, and tail; it rests in the upper abdomen and is connected to the spleen, stomach, duodenum and colon[1,2]. The pancreas is both an exocrine and endocrine gland with three major cell types: ductal epithelium, acinar, and islet cells[2].

The endocrine pancreas consists of the islets of Langerhans and functions primarily in regulation of blood glucose. The islets of an adult are comprised of four cells types: alpha, beta, delta, and pancreatic polypeptide cells, which produce glucagon, insulin, somatostatin, and pancreatic polypeptide, respectively[3]. Islet cells are spherically arranged with Beta cells at the centre surrounded by alpha and delta cells. The hormones produced by islets are released directly into the blood stream to maintain glucose homeostasis[4].

The majority of pancreatic mass is the exocrine pancreas; the ductal epithelial and acinar cells[3]. The acinar cells comprise 82% of pancreatic volume and produce several proenzymes or inactive precursors of enzymes which are transported to the duodenum through the pancreatic duct to aid in digestion. Amylase, lipase, and protease are the three types of digestive enzymes acinar cells produce[4]. Acinar cells form lobules connected to branching tubules of the intercellular canaliculus which is made up of duct cells[2]. The secretions of the acinar cells flow into the intercellular canaliculus along with the sodium bicarbonate secreted by duct cells, this branching system eventually converges into a main collecting duct, which then joins with the common bile duct[4]. The most common pancreatic malignancy, pancreatic ductal adenocarcinoma (PDAC), arises from the ductal epithelial cells.

1.1.2 Transition from Precursor Lesions to Malignancy

There are three common types of non-invasive precursor lesions of which PDAC can arise: pancreatic intraepithelial neoplasia (PanIN), intraductal papillary mucinous neoplasm (IPMN), and mucinous cystic neoplasm (MCN)[5]. IPMN and MCN progress from

low grade dysplasia, through intermediate dysplasia, to high grade dysplasia, whereas PanINs progression is classified into 1A, 1B, 2, and 3 lesions[5].

There are two modes of PDAC development, the first is a stepwise progression of precursor lesions into PDAC. The progression starts with telomere shortening, which causes chromosomes to become unstable and can lead to abnormal chromosome fusions[5]. Activating mutations in oncogene KRAS, as well as hypermethylation of CpG islands can be found in lower grade lesions[5,6]. The progression continues in higher grade lesions and invasive carcinomas, in which inactivating mutations in tumour suppressors genes (TSG) *CDKN2A*, *TP53*, and *SMAD4* are found[5,6]. Nearly all PDACs have activating *KRAS* mutations, while the frequency of mutations in *TP53*, *CDKN2a* and *SMAD4* have been reported at 74%, 31% and 35%, respectively[7]. PDACs harbour high levels or structural rearrangement and one study found that 45% of PDACs display polyploidization[7,8]. The second mode of PDAC development is a chromothripsis-driven development; a single catastrophic genomic event that can produce multiple driver events simultaneously and lead to neoplastic transformation[9,10].

1.1.3 Pancreatic Ductal Adenocarcinoma

Pancreatic cancer is currently the third leading cause of cancer death in Canada, with a 5-year survival rate of less than 10%[11]. The low survival has been attributed to the aggressive nature of the disease, with the majority of patients presenting at advanced stages, and the lack of therapeutic options[12]. The only curative treatment is surgery, and fewer than 20% of patients diagnosed with PDAC are eligible. Even after surgery, the 5-year survival rate remains low, at approximately 10% without adjuvant chemotherapy and between 15 to 20% with adjuvant chemotherapy, due to metastatic disease or locoregional recurrences[12,13].

The majority of patients present with unresectable tumours either locally advanced disease, non-metastatic but unresectable disease due to vascular involvement, or metastatic disease[14]. The standard of care and first-line systemic therapy for locally advanced or metastatic PDAC includes Gemcitabine, either as a monotherapy or with additives such as albumin-bound paclitaxel or erlotinib, and FOLFIRINOX[14–16]. Gemcitabine has been accepted as standard first line therapy since 1997; an anti-

metabolite, nucleoside analog that interferes with DNA synthesis[17,18]. FOLFIRINOX is the combination of 5-fluorouracil (5FU), leucovorin calcium, oxaliplatin and irinotecan and has shown survival advantage compared to Gemcitabine, but increased toxicity was reported by Conroy, et al[19,20]. Similar to Gemcitabine, 5FU is an anti-metabolite, leucovorin stabilizes the complex 5FU forms to inhibit the formation of thymidine, which in turn inhibits DNA synthesis. Oxaliplatin and irinotecan both cause different types of DNA damage which ultimately lead to apoptosis[19].

With improvements in sequencing technology and a movement towards precision oncology, several targeted therapies have been explored in PDAC. PDAC subtype classification systems have been published, most using large scale genomic analyses. The collision classification used transcriptional profiles of primary PDAC samples, as well as human and mouse PDAC cell lines to define three subtypes: classical, quasi-mesenchymal (QM), and exocrine-like[21]. Collission and colleagues found that these subtypes could be used as a predictor to overall survival after resection, independent of grade or stage[21]. Moreover, they found the QM subtype to be more sensitive to gemcitabine than the classical subtype, while the classical subtype was more sensitive to erlotinib, attributing these differences in part to the differences in KRAS expression between the subtypes[21]. The Moffit classification also used transcriptional profiling to define their subtypes, but included stroma subtypes as well as tumour subtypes, defining activated and normal stroma and classical and basal-like tumours[22]. Basal-like tumours with activated stroma were found to have the worst prognosis, followed by basal-like with normal stroma, and the classical tumours[22]. The Bailey classification identified four subtypes: squamous, pancreatic progenitor, immunogenic and aberrantly differentiated endocrine exocrine, they again used transcriptional information and associated these subtypes with histological features[23]. All the subtypes have unique insights and can inform about prognosis and have potential for clinical use.

Although molecular subtyping cannot yet inform treatment decisions, genetic testing has led to investigation of targeted therapy in PDAC. A trial for patients with mutations in genes of the homologous directed repair pathway (HDR), specifically *BRCA1* and *BRCA2*, treated with poly adenosine disphosphate-ribose polymerase inhibitor Olaparib as maintenance therapy was shown to increase progression-free survival

compared to placebo[24]. Patients harbouring mutations in the DNA mismatch repair (MMR) pathway have also received targeted therapies on trials evaluating the efficacy of programmed cell death-1 inhibitor pembrolizumab. One trial, which evaluated several noncolorectal solid tumours for their response to pembrolizumab, found pembrolizumab to be beneficial in 18.2% of patients with MMR deficiency[25]. Another trial, which focused solely on pancreatic cancer, found treatment with pembrolizumab to be beneficial in 4 of their 5 patients with MMR deficiency[26].

1.2 Hereditary Pancreatic cancer

1.2.1 Genetics of Hereditary PDAC

Up to 10% of incident PDAC cases are considered to be hereditary, with the patient having a first-degree relative (FDR) that had also been affected by PDAC. Hereditary cases can fall into two categories; those with a known cancer syndrome gene and those without. Individuals without a known mutation, but with a strong family history of PDAC are considered to have familial pancreatic cancer (FPC).

The highest risk predisposition syndrome is Peutz-Jeghers Syndrome (PJ), the cumulative risk has been reported at 36% from age 15 to 64[27]. PJ is an autosomal dominant-inherited, polyposis disorder caused by mutations in *STK11*, characterized by poylps throughout the gastrointestinal tract[27]. PJ increases cancer risk of several different cancers, not just PDAC, with a lifetime risk of 93% of developing cancer in individuals afflicted[27]. *CDKN2A*, a frequently mutated gene in sporadic PDAC, is the causal mutation in an estimated 40% of Familial Atypical Multiple Mole Melanoma syndrome (FAMMM). Individuals with FAMMM caused by *CDKN2A*, but not other mutations, have an increased risk of developing PDAC with a lifetime risk of around 17%[28,29]. Individuals with mutations in *PRSS1* leading to hereditary pancreatitis have a cumulative risk of nearly 40% of developing PDAC and the probability of developing PDAC increases in individuals with a paternal inheritance pattern[30]. The most commonly mutated susceptibility genes in pancreatic cancer are involved in hereditary breast and ovarian cancer syndrome (HBOC): *ATM* and *BRCA2*[31]. Both *ATM* and *BRCA2* are involved in the double strand break repair pathway and deleterious mutations in either of these genes are found in approximately 2% of PDAC patients. *BRCA1* and *PALB2* are also implicated HBOC and

increase risk to PDAC, however they are found at a lower frequency in PDAC patients at 0.5 and 0.4%, respectively[31] .

FPC accounts for 85-90% of all hereditary PDAC cases. FPC describes a family with no known causative genetic mutation and at least two FDR with pancreatic cancer[32]. It has been reported that individuals with two FDRs affected by PDAC have a 6-fold increased risk of developing it themselves, and this risk increases with a greater number of affected relatives[33]. Little clinical difference has been found between sporadic and familial PDAC, unlike other cancer syndromes, early age of onset is not a hallmark of FPC though some studies have found an earlier onset, and no difference in overall survival has been observed, but FPC are more likely to have multifocal precursor legions[34]. In order to investigate whether there was a missing major factor influencing susceptibility to PDAC, Klein and colleagues performed segregation analysis on 287 families to determine whether the pattern of PDAC in families was consistent with genetic transmission[35]. The modelling in this study provided evidence for a rare major gene that influences age-at-onset of PDAC and underscores the importance of the search for more PDAC predisposition genes[35].

1.2.2 Surveillance for individuals at high risk for PDAC

A strategy to improve survival for PDAC is surveillance of at-risk individuals for detection of early stage and potentially curable PDACs as well as precursor lesions with potential for malignant transformation (i.e., PanINs and IPMNs). A surveillance study that recruited 354 high risk individuals over 16 years and screened individuals with endoscopic ultrasound, magnetic resonance cholangiopancreatography, or computed tomography at intervals determined by the presence or absence of pancreatic lesions[36]. The individuals diagnosed with PDAC while under surveillance survived an average of 5.3 years, significantly longer than those who developed PDAC after stopping surveillance, 1.4 years[36]. Ninety-three percent of patients diagnosed with PDAC on this study were not known mutation carriers and only had a family history of PDAC[36]. A surveillance study across three different European centres found surveillance to be most beneficial to *CDKN2A* mutation carriers, of which 7.3% were diagnosed with PDAC and 75% of patients diagnosed were able to undergo resection[37]. Though patients with other

mutations (such as *BRCA2*) or FPC were also diagnosed over the course of this surveillance study, the most benefit was shown for *CDKN2A* mutation carriers[37]. Shorter term surveillance studies have identified lesions, cysts, or dilated ducts, and in some cases surgically intervened, however the authors could not conclude the benefits as to PDAC survival[38,39].

1.3 *SMG1* as a PDAC predisposition gene

1.3.1 Identification of *SMG1* as a PDAC predisposition gene

With the goal of uncovering missing genetic predisposition in FPC, the Zogopoulos lab previously published a case-control region-based gene association study that identified *SMG1* as a candidate PDAC susceptibility gene[40]. This study included a discovery series with one cohort of FPC and young onset cases, a second cohort of 289 unselected pancreatic cancer cases, and a control group of 987 DNA samples from individuals with no personal history of cancer. Variants in 445 DNA damage response and repair genes were assessed for an association with PC risk using the Mixed Effects Score Test (MiST) and Leave-one-out sensitivity analysis. MiST tests the association of genetic variants in a region and a phenotype and incorporates additional information about variants, including the Combined Annotation Depletion Dependent score and the type of mutation. Forty-eight genes were found to have a p-value of <0.05 after MiST analysis, but only *RECQL, CHEK2, TDG, SMG1, BRAC1, BRCA2*, and *STK11* remained significant after correcting for multiple comparisons.

Complementing the MiST analysis, Leave One Out (LOO) analysis was performed to identify which variants were driving association. LOO analysis consists of two tests: the first of which splits a gene into windows of 30 variants with windows overlapping with at least 10 variants, each window is dropped one at a time, and the MiST p-value is recalculated. A p-value increase suggests at least on risk variant resides within the dropped window. The second LOO test sequentially dropped each variant within a window found to contain a risk variant. MiST p-values were recalculated after each variant was dropped and a variant that increase the p value by ≥35% was classified a candidate pathogenic variant. *SMG1* was the only novel gene found to contain variants driving association in the case series, rather than the control series. Fourteen unique variants in

27 cases and 2 controls were found to drive association with PC risk. Findings were validated using an independent case-control series of 532 FPC cases and 753 non-cancer controls, which found *SMG1* variants in 41 FPC cases and 32 controls.

1.3.2 PI3K-Related Kinase Protein Family Structure

Suppressor of morphogenesis in genitalia-1, or *SMG1* nonsense mediated mRNA decay associated PI3k related kinase (*SMG1*), is the most recently discovered member of the phosphatidylinositol-3 related kinase (PIKK) family of serine threonine kinases[41]. *SMG1* is one of six mammalian PIKKs, other family members include *mTOR*, *ATR*, *ATM*, TRAAP, and DNA-PKcs[42]. PIKKS are a physically large family, ranging in size from 280 to 470KDa, SMG1 is on the larger end at 410kDA[43,44].

There are three common domains across PIKKs: the FRAP-ATM-TRRAP (FAT) domain, the PIKK kinase domain, and the FAT-C terminal (FATC) domain[43,45]. The N-terminus of PIKKs consists of alpha helices, the majority of which assume a tertiary structure resembling Huntingtin Elongation factor 3, Alpha-regulatory subunit of protein phosphatase 2A and TOR1 (HEAT) repeats[44]. Following the N-terminal is the FAT domain, adjacent to the FAT domain is the PIKK kinase domain, followed by the FATC domain on the C-terminus. The FATC and FAT domains are thought to interact, mediating protein-protein interaction and ensuring proper functioning of the kinase domain. Unique to SMG1 is a ~1000 amino acid insertion between the PIKK and FATC domain, in other PIKKs this area is 16 to 82 amino acids long[45]. The FATC domain is a small, 30 amino acid domain and is highly conserved among PIKKs. One study investigating the functional consequences of site directed mutagenesis in *SMG1* found missense mutations causing a change in amino acid sequence in the FATC domain could greatly reduce the kinase activity of SMG1; one mutation reducing the kinase activity to 7.9% of the activity seen in wild type SMG1, while a more conservative mutation reduced kinase activity by 50%[46].

1.3.3 *SMG1* and cancer

While *SMG1* as a predisposition gene is unique to PDAC, *SMG1* has been implicated in a number of cancers, including: gastric cancer, hepatocellular carcinoma, acute myeloid leukemia (AML), and head and neck squamous cell carcinoma (HNSCC)[47–

[50]. There are two mechanisms described to dysregulate *SMG1* expression in cancer. The first mechanism is downregulation of *SMG1* by non-coding RNAs, which has been observed in gastric, nasopharyngeal, and liver cancer and the second is hypermethylation of its promoter, observed in AML and HPV positive head and neck squamous cell carcinoma[49,50].

In gastric cancer, *SMG1* has been shown to be suppressed by miR-192 and -215. *SMG1* had significantly lower expression in tumour tissue than normal tissue and expression of *SMG1* was found to negatively correlate with serosal invasion and tumour size. Further, the inhibition of *SMG1* by miR-192 and -215 promoted cell proliferation, invasion, and migration, possibly through upregulation of downstream targets of the *Wnt* pathway [51]. Similarly, in nasopharyngeal carcinoma, overexpression of mir-18a has been observed in clinical samples and *SMG1* has been identified as one of mir-18a's target transcripts. In paired normal-tumour tissues, *SMG1* was found to have significantly lower levels of mRNA expressed in nasopharyngeal carcinoma samples than normal tissues. Cell lines overexpressing mir-18a were found to have increased invasion, proliferation, and migration, and these oncogenic effects were confirmed experimentally to be exerted via suppression of *SMG1* through use of siRNAs[52]. In hepatocellular carcinoma, the mRNA expression and protein levels of SMG1 were found to be lower than in patient-matched normal samples and the lower levels of SMG1 were associated with poor tumour differentiation, advanced clinical stage, and poor overall survival[53]. More recently, a mechanism by which *SMG1* is suppressed in hepatocellular carcinoma was described; lncRNA MAGI2-AS3 was found to be significantly downregulated in HCC tissues compared to matched normal samples, as was *SMG1*. MAGI2-AS3 acts as molecular sponge to regulate miR-374b-5p, an miRNA shown to target and downregulate *SMG1*. Overexpression of MAGI2-AS3 correlated with increased expression of *SMG1* as well as inhibited cell proliferation and migration[48].

The second mechanism of *SMG1* dysregulation in cancer is epigenetic dysregulation. Epigenetic dysregulation has been reported in AML and HPV positive head and neck cancer (HPPHNSCC). In HPPHNSCC, the promoter of *SMG1* has been found to be hypermethylated, this process, controlled by E6 and E7 proteins of the HPV virus, results in lower expression of *SMG1*. In contrast to hepatocellular carcinoma, low *SMG1*

status in HPHNSCC correlates with improved overall and recurrence-free survival, likely due to the use of ionizing radiation in HNSCC treatment[50]. Cells expressing lower levels of *SMG1* were more sensitive to increasing doses of radiation in comparison to cell lines with high *SMG1* expression, defining the importance of *SMG1* as a determinant to radiation sensitivity. As in HPPHNSCC, *SMG1* has also been found to be downregulated in AML via promoter hypermethylation[54]. *SMG1* was found to have lower mRNA expression in AML patients than healthy controls and the majority of AML patients tested were found to have *SMG1* promoter hypermethylation, whereas no SMG1 hypermethylation was reported in controls[49]. The treatment of an AML cell line with decitabine, a demethylating agent, increased the expression of *SMG1*, providing evidence that promoter hypermethylation is the mechanism of suppression of *SMG1*[49]. Similar to cells with *SMG1* promoter methylation, cells transfected with siRNAs against *SMG1* were found to have fewer cells undergoing apoptosis, suggesting *SMG1* as a TSG in AML[49].

Among the best evidence for *SMG1* in cancer predisposition is from a mouse model investigating the physiological role of *Smg1* and the impact of *Smg1* deficiency through a heterozygous of *Smg1* genetrap ($Smg1^{gt}$) model. This study crossed two $Smg1^{gt/wt}$ mice to generate $Smg1^{gt}$ embryos and determined that Smg1 is required for murine embryogenesis; the authors found that Smg1 deficient embryos arrest at developmental stage E8.5[55]. $Smg1^{gt/wt}$ mice had a significantly reduced lifespan compared to normal littermates, and were prone to chronic inflammation, enlarged organs, and cancer. $Smg1^{gt/wt}$ mice developed significantly more cases of hematopoietic and lung cancers than normal littermates. The mechanism of tumourigenesis was speculated to be low-level inflammation due to elevated reactive oxygen or nitrogen species. Notably, NMD and stress response, two of *Smg1*'s best characterized cellular functions, were normal in $Smg1^{gt/wt}$ mice and dysregulation in these pathways was not implicated in tumourigenesis. More recently, a group led by Roberts, published a study on mice lacking one *Smg1* allele and completely deficient in Atm. The lifespans of the $Atm^{-/-}$ $Smg1^{gt/wt}$ were significantly shorter than both the $Atm^{-/-}$ and $Smg1^{gt/wt}$ genotypes alone. Evidence of blood cancer was found in all $Atm^{-/-}$ $Smg1^{gt/wt}$ mice, and some of these mice also had abnormal retention of fat, chronic inflammation, or extramedullary

haematopoiesis. The type of cancers found in *Atm*$^{-/-}$ *Smg1*$^{gt/wt}$ mice are the same as cancer in *Atm* null mice, the knockout of one copy of *Smg1* appears to increase the speed of tumourigenesis in *Atm* null mice. *Atm*$^{-/-}$ *Smg1*$^{gt/wt}$ mice exhibited higher levels of oxidative damage than found in either genotype separately, and mice lacking *Atm*, especially when combined with *Smg1* gene trap, are more likely to express higher levels of certain cytokines[56]. Depletion of *SMG1* in xenografts has also been shown to promote tumour growth, one studying reporting decreased time to reach tumour volume of 300mm^3 after implantation of xenografts expressing *SMG1* shRNA compared to an empty vector[57].

1.3.4 SMG1 Functions

SMG1 has been associated with several cellular processes, the overarching theme of these processes being genome and protein integrity. Human *SMG1* was identified simultaneously by two different groups, Denning and colleagues were the first to publish their findings and had discovered *SMG1* as a human ortholog of *C. elegans SMG1*[58]. *SMG1*'s *C.elegans* ortholog had been implicated NMD, which guided authors towards characterizing the kinase activity of SMG1 and lead to further studies confirming its role in NMD[58,59].

The second group to publish on human *SMG1*, Brumbaugh and colleagues, identified *SMG1* through its homology to other members of the PIKK family[41]. The role of genome integrity of other members of the PIKK family lead this group to characterize *SMG1* for roles in genotoxic stress response[41].

1.3.4.1 Nonsense mediated decay

NMD is a translation-dependent, eukaryotic mRNA surveillance pathway that degrades mRNAs with premature termination codons (PTC) due to mutation, transcription errors, or incorrect splicing[60]. The primary function of NMD is thought to be prevention of truncated proteins, though there is evidence that it may be involved in post-transcription gene expression. Some RNA binding proteins have been shown to use alternative splicing coupled with NMD as a feedback loop to regulate their expression. One study found that mRNAs encoding multiple snoRNAs in their introns were enriched for

nonsense splicing isoforms, and speculated that degredation through alternative splicing coupled with NMD could be a way of regulating snoRNA levels[61]. The concept of alternative splicing and NMD as means of gene expression was further supported in studies showing alternatively spliced ultra-conserved elements, which cause premature stop codons, in splicing regulator transcripts[62,63].

The current model of NMD in mammalian cells in the exon junction complex (EJC) model. NMD degrades PTC-harbouring mRNAs in the pioneer round of translation when a PTC is located 50-55 nucleotides upstream of an EJC[64]. It is not well understood how NMD-activating translation termination is triggered, but it is known that termination requires eukaryotic release factor and eRF3[64,65]. The exact order of complex binding and formation is not clear, but the favoured model suggests that UPF1 is recruited to the terminating ribosome through a direct interaction with eRF3, an interaction that is avoided during efficient translation termination[66,67]. Inefficient termination of translation or a ribosome stalled on a premature termination codon allows UPF1 to bind with eRF3, and also recruits SMG1c to interact with erF3-erF1 and UPF1 to form the SURF complex. SMG1c consists of SMG1 and its regulators SMG8 and SMG9[67]. If an EJC is located 50 or more nucleotides downstream of the termination complex, Upf1 can associate with Upf2, Upf3b and the EJC, this association between the SURF complex and EJC causes the formation of the decay-inducing (DECID) complex and promotes SMG1 to phosphorylate UPF1[67]. After phosphorylation, UPF1 associates with SMG5, SMG6, SMG7 and other general mRNA degredation factors. SMG6 initiates RNA degredation by cleaving the NMD target, while UPF1 associates with SMG5 and SMG7 to recruit POP2. The formation of SMG5, SMG7, UPF1 and POP2 initiates decapping and XRN1-mediated 5'-3' degredation[67].

Direct evidence of SMG1's involvement in NMD was shown shortly after the identification of human *SMG1*. Yamashita et al. showed that SMG1 directly phosphorylates UPF1 and kinase-deficient mutated SMG1 is unable to phosphorylate UPF1[59]. Further, they found that SMG1 interacts with other members of the surveillance complex (UPF1, UPF2 and UPF3), and inhibitors of SMG1 stabilized an endogenous PTC-containing transcript[59].

1.3.4.2 Genotoxic Stress Response

In addition to its activity in NMD, *SMG1* has been described to respond to genotoxic stress, specifically to ionizing radiation (IR) and Ultra Violet light (UV). The first study that identified the role of *SMG1* in stress response found that SMG1 exhibits p53 kinase activity, and siRNAs against *SMG1* or transfecting catalytically dead S*MG1* decreased P53 phosphorylation and accumulation after exposure to IR[41]. Further, this study looked at α-phospho-histone H2AX (γH2AX) phosphorylation, a target of response to DNA strand break in cells depleted of *SMG1* and found the siRNA against *SMG1* to drastically increase the number of γH2AX-positive nuclei, leading the authors to suggest that *SMG1*-deficient cells develop spontaneous DNA damage. After exposure to IR, cells transfected with antisense RNA against *SMG1* were found to have more cells arrested in the G2/M phase, which was attributed to a compromised G1 checkpoint[41]. Further, depletion of *SMG1* was shown to decrease survival after exposure to IR through clonogenic survival assay[41]. A later study replicated these results – *SMG1* knockdown depletes ser15-p53 levels after exposure to IR, and further showed that *SMG1* contributes to regulation of the G1/S cell cycle checkpoint in both a p53-dependent and independent manner[57]. Instead of cells arresting in the G1 phase after exposure IR causing an increase in the proportion of cells in G1 and preventing progression of cells into the S phase when DNA is replicated, cells depleted of SMG1 had a significant reduction in the number of cells in G1[57]. This group further implicated *SMG1* in the regulation of the G1/S phase checkpoint by uncovering SMG1's ability to phosphorylate Cdc25a for degredation in addition to P53 phosphorylation for stabilization and accumulation. Cdc25a removes inhibitory phosphorylation of CDK2, allowing it to activate progression in the cell cycle[68].

Another stimulus of genotoxic stress in which *SMG1* responds is hyperoxia. Hyperoxia is a clastogenic agent that cause double strand breaks; P53 responds to double strand breaks after its stabilization and phosphorylation[69]. Through knockdown by siRNAs, SMG1 was found to be required for the phosphorylation and stabilization of P53 after exposure to hypoxia, especially in the first 24 hours post-exposure[69,70]. Phosphorylation of P53 after exposure to hyperoxia is necessary for G1 checkpoint

activation, and in the first 24 hours of hyperoxia knockdown of *SMG1* diminished G1 checkpoint activation[71].

1.3.4.3 Stress granule formation

While the role of *SMG1* in NMD has been well characterized and there have been numerous studies on SMG1's P53-mediated stress response activity, there have been singular studies describing other functions of *SMG1*. One such study is the association of *SMG1* with stress granules. Stress granules are transiently formed, accumulations of mRNA and associated proteins that form in response to cellular stresses, such as oxidative stress and heat shock, which cause the activation of eukaryotic translation initiation factor 2 alpha[72]. SMG1 was found in stress granules formed after exposure to sodium arsenite (NaAs), which causes oxidative stress, after heat shock, and after exposure to H2O2, which causes DNA damage. Knockdown of *SMG1* was shown to reduce stress granule formation in cells treated with H2O2, but not in cells treated with heat nor with NaAs[72]. Overexpression of a kinase-dead SMG1 mutant did not reduce stress granule formation after treatment with H2O2 or sodium arsenite, and the authors concluded that SMG1's presence, but not its kinase activity, was required for stress granule formation is response to H2O2 and NaAs stimuli[72].

1.3.4.4 Telomere maintenance

Another briefly characterized role for *SMG1* is in telomere maintenance. Telomeres ensure chromosome stability during replication and provide a cells with a cellular clock that determines how many more replications the cell can undergo. In a study that identified TERRA, transcripts from mammalian telomeres, SMG1 was found to be enriched at telomeres. SMG1 and other NMD factors, including UPF1 and UPF2, were found to be associated with TERRA at the end of chromosomes. Reduction of either SMG1 and UPF1 by shRNAs or siRNAs lead to an increase in telomere-free chromosome ends and telomeric fragments, and an increase in TERRA presence at the end of chromosomes[73]. In yeast, induced transcription from telomeres caused telomere shortening, which lead authors to conclude that *SMG1* and other factors coordinate TERRA and telomeric enzyme activities that prevent telomeric shortening[73].

1.4 CRISPR/Cas9 Gene Editing

1.4.1 Gene editing and CRISPR origin

CRISPR/Cas9 (Clustered regularly interspaced short palindromic repeats) technology has started a new era in gene editing; CRISPR is relatively efficient, faster, easier, and cheaper than prior technologies[74]. The two prominent gene editing technologies preceding CRISPR were Zinc Finger Nuclease (ZFN) and Transcription-Activator Like Effector Nucleases (TALEN) editing. ZFNs consist of two DNA binding domains, two spacer regions, and a nuclease domain, consisting of the *FokI* restriction endonuclease, that is bound to either DNA binding domain by the spacer region. The DNA binding domains have three sets of zinc finger motifs, each zinc finger motif has the ability to bind to 3bp of DNA[75,76]. The ZFN forms a dimer and *FokI* cleaves each strand of DNA bound to the zinc finger, introducing a double strand break in the each spacer region[75]. TALENs also utilize the *FokI* endonuclease connected to two DNA binding domains. The difference between these systems lies in the DNA binding domains; TALEN DNA binding domains consist of conserved repeat sequences from transcription activator-like effector, a protein originally found in bacteria. A central region of the TALE protein contains specific sequence motifs allows TALE proteins to bind to DNA, and specificity is determined by repeat variable di-residues; two amino acids that differ in each repeat, the amino acid corresponding to specific nucleotides[77]. CRISPR/Cas9 technology has been adapted from the immune system of bacteria and archaea for genome editing. The CRISPR locus found in some bacteria and archaea contain integrated foreign DNA (spacers) separated by repeat sequences, hence the name clustered regularly interspaced short palindromic repeats[78]. Spacers are integrated based on the presence of a protospacer adjacent motif (PAM) sequence; they are transcribed into crRNA and join with tracrRNA to form a guide RNA[78]. The guide RNA associated with CRISPR associated protein 9 (Cas9), which introduces a double strand break in the DNA complementary to the guide RNA sequence, 3 bp upstream of the PAM sequence[79]. This system has been adapted for genome editing by combining the tracr and cr RNAs into a single guide RNA (sgRNA) with Cas9 from *Streptococcus pyogenes*, which recognizes a PAM sequence of 5'NGG-3'[79].

Each of the above-mentioned gene editing technologies introduce double strand breaks to target sequences. Depending on the DNA repair mechanism, this can result in the generation of indels, which usually result in a premature termination codon or, in the presence of a homologous DNA template, the targeted knock-in of a desired mutation. The introduction of a double strand break by Cas9 or *FokI* endonuclease predominantly triggers one of two DNA repair pathways in eukaryotic cells: non-homologous end joining (NHEJ) or homology-directed repair (HDR)[80]. NHEJ is the most common repair mechanism and essentially joins two pieces of DNA together, often introducing indels in the process[80]. HDR occurs when a homologous DNA template is present, the DNA can be repaired through recombination with the template, which also allows for the knock-in of mutations[80]. CRISPR has gained popularity over previous methods due to the inexpensive and ease of engineering components, simple cloning and oligonucleotide synthesis in comparison to expensive and complex molecular cloning or protein engineering[77].

1.4.2 CRISPR advancements

CRISPR has changed gene editing as we know it and following CRIPSR came many inventive modifications of the technology. Modification of the Cas9 enzyme to render it catalytically inactive or partially active has given rise to CRISPRi and CRISPRa, and nickase function, respectively. CRISPRi, or CRISPR interference, consists of catalytically dead Cas9 (dCas9), and sgRNA targeting a promoter region, which causes dCas9 to sterically block the association between cis-acting DNA motifs and transcription factors, ultimately leading to repressed transcription[81]. CRISPR activation (CRISPRa) also uses dCas9, but dCas9 is fused with transcription activator domains in order to upregulate gene expression from the sgRNA-target locus[82]. Another modified Cas9 mutant D10A Cas9 introduces single strand nicks instead of double strand breaks, and has been used with pairs of sgRNAs to target genomic loci for knockout[83]. This strategy decreases off-target effects as single strand breaks are repaired with high fidelity mechanisms; only the locus with two nicks, resulting in a double strand break, is likely to acquire indels[83].

The persisting difficulty of generating a targeted knock-in mutation has led to further innovation in CRISPR, either in optimizing the original components of CRISPR, adding compounds to traditional CRISPR, or modifying the underlying system. Several studies have increased efficiency of the knock-in by redesigning the DNA donor templates used. One of the simplest modifications was shown in Richardson et al, in which an asymmetrical donor DNA template was designed such that the majority of the template annealed with the PAM-proximal side of the double strand break and the donor template was complementary to the non-target strand[84]. The authors took into account the interaction between Cas9 and the target DNA, finding that Cas9 dissociated asymmetrically, releasing the 3' end of cleaved nontarget DNA strand first, and achieved a knock-in rate of 57%[84]. Covalently tethering the DNA template to the Cas9/sgRNA complex has been shown to increase the HDR frequency 30-fold, but requires a more complicated RNP transfection[85]. Co-CRISPR is a technique that was developed from the observation that simultaneous CRISPR-Cas9 mediated gene editing events at two unlinked loci are not statistically independent, which was observed in *Drosophila*, *C. elegans* and mammalian cells[86,87]. Co-CRISPR knocks in a visible phenotypic marker or mutation with the phenotypically silent mutation of interest[87]. A protocol for Co-CRISPR in human cells was developed using a sgRNA and donor template that knocks-in a point mutation in the Na+/K+ ATPase, which gives resistance against the drug ouabain[87].

HDR is restricted to certain points in the cell cycle and is less likely to be the pathway utilized for DNA repair, as such, individuals have optimized knock-in timing with respect to cell cycle or added compounds to inhibit NHEJ or activate HDR. One study focused on enhancing HDR by co-expressing proteins in the HDR pathway, and ultimately fused *Rad52* with Cas9; co-expression of HDR activator *Rad52* increased knock-in efficiency 2.8 fold over the control group and the Cas9-Rad52 fusion increased knock-in efficiency by 3.4 fold[88]. In order to suppress NEHJ, inhibitor of DNA ligase IV Scr7 has been described to significantly increase the frequency of HDR-enabled knock-in[80]. The use of inhibitors that arrest cells at specific phases of the cell cycle in which HDR is active has been reported to increase efficiency of knock-in 3 fold[89].

Systems that circumvent the need for HDR altogether include homology-independent target integration (HITI), base editing, and prime editing. HITI uses circular

donor templates that harbour the sgRNA target site and are integrated through NHEJ after Cas9 introduces double strand breaks both in the genome and the DNA template[90]. HITI has demonstrated highly efficient insertion of donor DNA, with applications in gene correction, but it is not suitable for the introduction of point mutations[90]. Base editing adds cytidine deaminase to the CRISPR/Cas9 system, utilizing a catalytically dead Cas9 fused to cytidine deaminase to introduce targeted C→T or G→A mutations[91]. Base editing addressed the need for improvements to CRISPR for point-mutation knock-in and reduced the frequency of indels due to the elimination of double strand cleavage, but has limitations in the target window and the types of mutations it can introduce. The biggest and most recent stride forward in increasing the efficiency of knocking-in point mutations is prime editing. Prime editing uses a Cas9 (h840A) nickase fused to reverse transcriptase and a prime editing gRNA (pegRNA) that has a primer binding site and the reverse transcriptase template that includes the edit[92]. The prime editing system combines the DNA template and the sgRNA in one, and increased the efficiency of knocking-in point mutations significantly, over 60% at some loci, in comparison to a maximum of 20% using the traditional HDR approach[92].

1.4.3 Translational applications of CRISPR

CRISPR is being used to study a broad range of human diseases; metabolic diseases, neurodegenerative diseases, cardiovascular diseases, and cancer. CRISPR has been important in cancer research to gain insight into chemotherapy sensitivities, modeling tumour evolution, and modeling variants of unknown significance (VUS)[93–98].

Most relevant to this dissertation is the use of CRISPR to model VUS. CRISPR has been used to model rare missense variants in *BRIP1* gene, the variants were identified on ClinVar and cell lines were generated using the traditional CRISPR knock-in approach, transfection of vectors with DNA template oligos with the addition of DNA ligase inhibitor SCR7 to prevent repair through NHEJ pathway[99]. The missense variants were further characterized for their sensitivity to intercross link damage as the *BRIP1* protein product acts in the double-strand break repair pathway[99]. CRISPR knock-in has also been used to evaluate missense VUS in *MSH2* genes, one of the causative genes of Lynch Syndrome, for their effects on DNA mismatch repair. This study had important

implications as VUS in causative Lynch Syndrome genes can have implications on diagnosis of Lynch and on treatments in cancer[96]. Similarly, the 6 most frequent *TP53* missense mutations of myelodysplastic syndromes were characterized in cell lines generated using CRISPR/Cas9 were determined to be loss-of-function with dominant negative effects[97]. Together these studies demonstrate the utility of CRISPR in modelling VUS associated with cancer and determining their functional significance.

1.4.4 Limitations of CRISPR

While CRISPR has advanced gene editing immeasurably, it is not without limitations. CRISPR has been shown to be advantageous over previous methods due to its high efficiency and low rates of off-target editing[100]. Knocking-in mutations, especially point mutations, remains a challenge in the field, due to the low HDR pathway efficiency in mammalian cells. As mentioned previously, research has been undertaken to address the low efficiency of knock-in mutations including advances such as base-editing, prime-editing, tethered donor templates, and the use of NHEJ inhibitors[80,85,91,92].

Beyond the latter biological limitations of CRISPR there are also technical considerations and model limitations. Technical considerations such as delivery of the CRISPR constructs depend on the model chosen. In cell lines, the success of CRISPR can depend on the efficiency of transfection, the type of molecule transfected, and whether the cells are able to grow from a single cell[101]. Cell survival and model success may also depend on the gene targeted and whether the cell can tolerate knock out or mutation of the gene of interest. Certain genes can be essential to cell survival and mutagenesis of these genes can result in cell death; the genes that are essential to cell survival can be different depending on the cell line[102]. CRISPR gene editing in cell lines is well-established technique, but modelling in cell lines has limitations. Cultured cells do not interact with different cell types, they are not part of a tissue architecture, and are not subject to an immune environment. Using CRISPR to model mutations in cell lines can provide insight into functional consequences of mutations of interest, but that insight may not translate at an organismal level[103].

1.5 Hypothesis and Rationale

To fine-tune surveillance protocols for individuals at high-risk of PDAC, a better understanding of the genetic landscape of PDAC predisposition is needed. A region-based gene association analysis from our lab previously identified *SMG1* as a PDAC predisposition gene and identified missense variants associated with PDAC. *SMG1* dysregulation has been previously associated with cancer, but there is no literature on germline variants in *SMG1* conferring risk, and the functional consequences of the variants identified in Wong et al, is unknown. We prioritized variants c.4249A>G and c.10921A>G identified in Wong et al. for functional characterization. As described in Wong et al, variant c.4249A>G segregated in three affected individuals in an FPC family. In previous SMG1 protein functional studies, the FATC domain has been found to be highly sensitive to mutagenesis with non-conservative missense mutations resulting in loss of up to 92% of function and conservative missense mutations resulting in up to 50% reduction of function. The missense variant c.10921A>G is a conservative mutation that the missense variant c.10921A>G is located in the FATC domain, which has been reported to be highly sensitive to mutagenesis. I hypothesized that *SMG1* acts as a TSG in PDAC and *SMG1* missense variants c.4249A>G and c.10921A>G are causal.

1.6 Specific Aims

1. To generate cell lines using CRISPR/Cas9 gene editing harbouring c.4249A>G and c.10921A>G missense *SMG1* variants.

2. To functionally characterize missense variants c.4249A>G and c.10921A>G cell lines for tumour-suppressor and SMG1-specific functions.

Chapter II: Materials and Methods

Cell culture

HEK293T and U2OS cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) (Wisent Bioproducts, 219-010) consisting of 4.5g/L glucose, L-Glutamine and sodium pyruvate, without bicarbonate. Medium was supplemented with 10% fetal bovine serum (FBS) (Wisent Bioproducts 080-150) and 1% antibiotic-antimycotic solution (Wisent Bioproducts 450-115). Cells were cultured to reach 70-80% confluency before subculturing. In each passage, cells were detached using 0.05% trypsin-edta (Wisent Bioproducts 325-042), following a 1X PBS rinse (Wisent Bioproducts 311-010).

HAP1 cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM) (Wisent Bioproducts 319-105) with L-Glutamine and hepes, supplemented with 10% FBS and 1% antimycotic solution.

CRISPR sgRNA plasmids and DNA Donors

Genomic locations of target sites were identified on UCSC genome browser build GRCh37 as chr16:18820956 and chr16:18866212 for c.10921A>G and c.4249A>G, respectively. sgRNAs for target sites were identified using webtool CHOPCHOP: the genomic location of each variant was entered into CHOPCHOP and sgRNAs selected based on the output are shown in Table 1. Oligonucleotides of the sgRNAs and antisense oligos were ordered with adapter sequences for restriction digest by *AgeI* and *SphI* and were cloned into LeGO/mU6-2/Cas9(NTD-FLG/NLS)-iG (Figure 1A). Template DNA was ordered from Genscript cloned into a pUC57 plasmid, the sequences of the template DNA are in Table 2 and the vector maps in Figure 1B. Template DNA corresponding to sgRNA E was cloned into the donor pUC57 plasmid corresponding to sgRNA D using a GBlock from IDT.

Table 1. sgRNA sequences and corresponding variant

Variant	sgRNA ID	sgRNA sequence
c.4249A>G	sgRNA A	ttagacctaattccatgaga
c.4249A>G	sgRNA B	ccatgagatggcttctaatt
c.4249A>G	sgRNA C	tccatgagatggcttctaat

c.10921A>G	sgRNA D	gtcaacctgtaaaaataggc
c.10921A>G	sgRNA E	gacaggttgactatgtcatta

Table 2. DNA template sequences

Variant	Corresponding sgRNA	Template dna sequence
c.4249A>G	sgRNA A and sgRNA B	aattctcctgcctcagcccctcaagcagccggaactacaggcatgttccaccac acccccctaatttctttgaatttcagtatagctgagatttcactctgttgccca ggctggtgtcaaactcctaagctcaaacaatcctcccatcttggcctcccaaag tgctcagattgtaagcataagcacctggcccaaaataataaagttaaaaaaaca tgtaaattcagtgaccaatgatctccggaaatacttcttactgtaatacattgc aacaaaattcttttgaaagcacagaattataaaggtaagtcagtaataagcaac ttattattcaggatgcaaatgttctgcaacagctggaaatgatcaactttgcta ttactaccttataccatacatacacagtaacattcatatacatgatttttatta tactcatgtattaaccataattaatttttactaacctgctgtataaagcaatttg gtttttttcaatatcaagttcgggcccccatttttcatccacttgaccttgggtt gatagtttttttaaaatgttggactaaatcctgtgcagtggtggtctttcccagc tgaacttcactgcactgtgccagcagtcttgttgcaagggacacattccctcgt tttctagcaaatttttgctgctgttagacctaattccatgagatgactttctaact ggaactgtttgttctgaagagaggaaaacgttttattagttcctaatttgtcta aaccagactgcatggaaacatggccaaaaaattattccccaactatccctcttc tgcaccacctcctgaaacaaagataactcatcctcaaatccctagcagctagca cagtgcttgagatacaatgaacaggttttgctaaattaaatgtatacaatgaag taaagtcaagtcattcatttcaagaaatttctttttagtttctgaatctttggc aaaaagcctgggccaacgggtagggaaagcaaggaagcagggaaggctaaacca tccccccgcacttcttctgtaagtactctctgcccttccaccattctcctggtt ttcttcctatcttgctgaagaatcctcagtctctttctcttccatcactcatcc attaaacaccagtgctccttggactcagtcctaggctctctacttttcccactg catactctccctggacatcatcatctttttcctgtgtcttcaattaccacagaaa tgctgacaactgctgaatcttcacccagacctgatatctgaatcccacctgcca actggtcatctacactctatttatttcaagagtgcttcctctaatgtcagtcat gccaaatggagctcattttccaccaccctccaaatcttttactccctttttt
c.10921A>G	sgRNA D	tctttctcagtattttacatcaagagacacatgtcagtttgtgccatctttggt ggtgataagctagatcatttggtcaagagaggtatcctcttgatgtcttcatta taaatgtacatttttttcctgtgggataaaaattaagagactgaatatcctgtt cttccatggtttggcatccgtaattcaatttttaacatcagcaaggaatttttt tttcttttaaaacactattatattcttgtataggacctagacttgaactcttctt gggcaagaatttttttttttttaagaaaccgttttgtcaagtttgtcaaaaacc atgagagattttctgagtttgggatatttttacttaagaaacgttctggtaggt taatggattgtgtgcttttcatatctttataactgtcttagctctctttttaaa gtatacttggttattgtgctttttattctgatagttattacatttttatatgaaa attaaggggcttacaggagcagggtaagtatagagtcctttgcacttaacttt ggcataatttgtctgctgattgtaattcagcagacaattgaaaagctttaaatg actttacatactatcatgcctatttttacaggttgactatgtcgttaaggaagc aactaatctagataacttggctcagctgtatgaaggttggacagcctgggtggg aagcggagctactaacttcagcctgctgaagcaggctggagacgtggaggagaa ccctggacctatggtgagcaagggcgaggagctgttcaccggggtggtgcccat cctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcga gggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccac cggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgt gcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtc cgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacgg caactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccg catcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaa gctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaa gaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgt gcagctcgccgaccactaccagcagaacacccccatcggcgacggccccgtgct gctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaa cgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcac tctcggcatggacgagctgtacaagtagtgaatggcaagacagtagatgagtct ggttaagcgaggtcagacatccaccagaatcaactcagcctcaggcatccaaag

		ccacaccacagtcggtggtgatgcaactgggggcttactctgaggaaacctagg aaatctcggtgcactaggaagtgaatcccgcaggacagctgcactcagggatac gcccaacaccatggcctgcaaccccagggtcaagggtgaaggaaagcaagctca ccgcctgaacacggagattgtctttctgccacagaacagcagcagacgtgtcgg gaggttagctgcggaaagaaatcgggatgccgcggagcacagagtgatttggaa ctccattccacctgaccctgtgtgtacaatccaggaaaaaaacaaaccccactc agaaacagagaaaactggggtcgcgaagaaatcacagccaaggaagatttgatg cattcagattctcgtgtaacacttgttgcttggcaacagtactggttgggttga ccagtaagtagaaaaaggctaaaggctatgcgatatgaatttcagaaatggact gaaaatggagagctatgtaacagatacactacagtagaagaacttacttctgaa atgaagggaaaaaaaccaccccatcgttccctactcctccccaccacttacccg ttcccccttacctaatctagtagat
c.10921A>G	sgRNA E	tctttctcagtattttacatcaagagacacatgtcagtttgtgccatctttggt ggtgataagctagatcatttggtcaagagaggtatcctcttgatgtcttcatta taaatgtacatttttttcctgtgggataaaaattaagagactgaatatcctgtt cttccatggtttggcatccgtaattcaatttttaacatcagcaaggaatttttt tttctttttaaacactattatattcttgtataggacctagacttgaactcttctt gggcaagaatttttttttttttaagaaaccgttttgtcaagtttgtcaaaaacc atgagagattttctgagtttgggatatttttacttaagaaacgttctggtaggt taatggattgtgtgcttttcatatctttataactgtcttagctctcttttttaaa gtatacttggttattgtgctttttattctgatagttattacattttatatgaaa attaaggggcttacaggagcagggtaagtatagagtcctttgcacttaacttttt ggcataatttgtctgctgattgtaattcagcagacaattgaaaagctttaaatg actttacatactatcatgcctatttttacaggttgattacgtggttaaggaagc aactaatctagataacttggctcagctgtatgaaggttggacagcctgggtggg aagcggagctactaacttcagcctgctgaagcaggctggagacgtggaggagaa ccctggacctatggtgagcaagggcgaggagctgttcaccggggtggtgcccat cctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcga gggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccac cggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctacggcgt gcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtc cgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacgg caactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccg catcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaa gctggagtacaactacaacagccacaacgtctatatcatggccgacaagcagaa gaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgt gcagctcgccgaccactaccagcagaacaccccatcggcgacggccccgtgct gctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaa cgagaagcgcgatcacatggtcctgctggagttcgtgaccgccgccgggatcac tctcggcatggacgagctgtacaagtag**tga**atggcaagacagtagatgagtct ggttaagcgaggtcagacatccaccagaatcaactcagcctcaggcatccaaag ccacaccacagtcggtggtgatgcaactgggggcttactctgaggaaacctagg aaatctcggtgcactaggaagtgaatcccgcaggacagctgcactcagggatac gcccaacaccatggcctgcaaccccagggtcaagggtgaaggaaagcaagctca ccgcctgaacacggagattgtctttctgccacagaacagcagcagacgtgtcgg gaggttagctgcggaaagaaatcgggatgccgcggagcacagagtgatttggaa ctccattccacctgaccctgtgtgtacaatccaggaaaaaaacaaaccccactc agaaacagagaaaactggggtcgcgaagaaatcacagccaaggaagatttgatg cattcagattctcgtgtaacacttgttgcttggcaacagtactggttgggttga ccagtaagtagaaaaaggctaaaggctatgcgatatgaatttcagaaatggact gaaaatggagagctatgtaacagatacactacagtagaagaacttacttctgaa atgaagggaaaaaaaccaccccatcgttccctactcctccccaccacttacccg ttcccccttacctaatctagtagat

Missense mutation

Modified to eliminate PAM sequence

P2A Sequence

Eukaryotic GFP

Stop codon

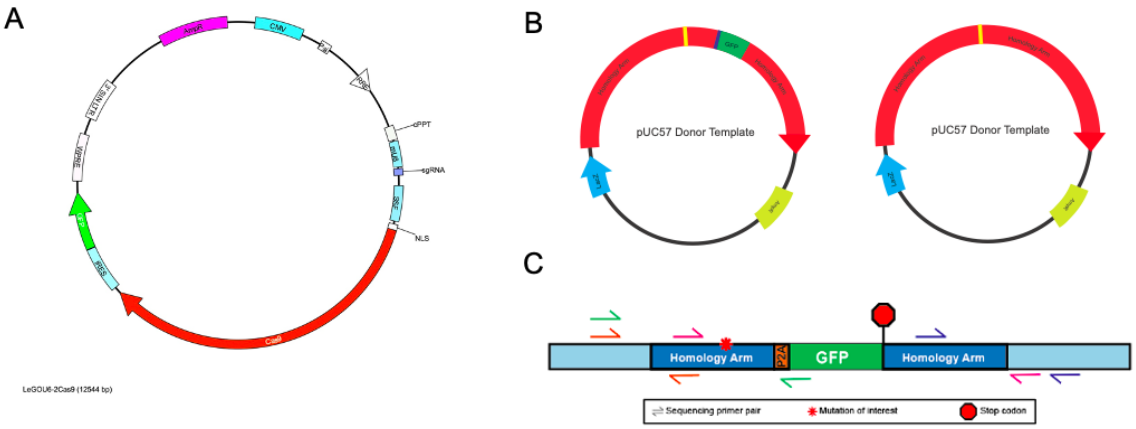

Figure 1. DNA template and sequencing primers for SMG1 c.10921A>G CRISPR knock-in. A) LeGOU6-2Cas9 vector map, c.10921A>G sgRNA-expressing vectors did not express GFP, but all other features were the same. B) pUC57 template DNA vector schematics for variant c.10921A>G and c.4249A>G. C)The c.10921A>G construct with a self-cleaving P2A sequence and GFP immediately adjacent to the SMG1 stop codon. Primer pairs for sequencing shown and forward and reverse pairs are colour-coded.

CRISPR transfection and clone screening

c.10921A>G Variant

HEK293T Cells were seeded in 6-well plates (8.0 x 10^5 cells per well) in complete DMEM and 24 hours later the plasmid donor template and LeGO-Cas9-sgRNA vector were co-transfected at a 3:1 ratio using Lipofectamine 3000 (Thermofisher, L3000). Cells transfected with sgRNA and DNA templates for variant c.10921A>G were cultured for one week and then single-cell sorted using FACS detecting GFP expression, as the donor DNA construct knocks-in GFP along with the *SMG1* missense variant (Figure 1C). U2OS cells were nucleofected using the Lonza Nucleofector Kit V (Lonza VCA-1003) according to manufacturer's protocols following the same workflow as HEK293T cells.

c.4249A>G Variant

HAP1 cells were seeded in 6-well plates (7.5 x 10^5 cells per well) in complete IMDM and were co-transfected 24 hours later with the plasmid donor template and LeGO-Cas9-sgRNA vector at a 3:1 ratio using Turbofectin 8.0 (Origene, TF81). Transfected cells were culture cultured for 48-72 hours and then single cell sorted by FACS; GFP-positive cells indicating successful transfection of the LeGO-Cas9-sgRNA vector that expresses GFP.

Cells were grown from single cells in 96-well plates to cell lines in 6-well plates, then screened for mutations using PCR.

Table 3. Clone screening and sequencing primers

Gene	Variant	Target Area	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Size (bp)
SMG1	c.10921A>G	5' End of homology arm	GTTCCTGAGCCTTTCCCTCTTC	CCAACCTTCATACAGCTGAG	846
SMG1	c.10921A>G	3' End of homology arm	TACTCTGAGGAAACCTAGGA	GCTATCACAGAAGAGGGAAA	892
*SMG1**	c.10921A>G	Targeted mutation site and 5' end of homology arm	GTTCCTGAGCCTTTCCCTCTTC	GTGAACAGCTCCTCGCCCTT	906
SMG1	c.4249A>G	Targeted mutation site	GGCCATGAGGGTTGAAAGTAAAG	TCCACTTGACCTTGGGTTGAT	975

*Selectively amplifies *SMG1* with c.10921A>G mutation and GFP insert

Polymerase chain reactions

To screen clone cell lines, DNA was isolated according to manufacturer protocol with gSYNC DNA Extraction kit (Geneaid GS100) and PCR reactions were run using Phusion Polymerase (Thermofisher, F530). Screening primers were designed such that at least one primer was outside the homology arms of the donor template. Regions amplified include region surrounding and including 5' homology arm, 3' homology arm, and mutation site (Table 3). All amplicons were sent for sanger sequencing and were aligned against the wild type amplicons and reference sequence using SnapGene.

MiSeq sequencing was performed using genomic DNA extracted using protocol described above. Amplicon for MiSeq was amplified using Phusion Polymerase (Thermofisher, F530) and sent for barcoding sequencing, barcodes added to the end of PCR primers (Table 4). Sequences were viewed and aligned on IGV using build GRCh37, build 19.

Copy number was quantified using qPCR on genomic DNA. DNA was extracted from 1 x 10^6 cells per cell line of the three clonal cell lines, wild type HEK 293T cell lines and diploid cell line MRC-5. qPCR was performed using PowerTrack SYBR Green Master Mix (ThermoFisher Scientific, A46012) on a Step One Plus Real Time qPCR machine (Applied Biosystems) data was quantified using StepOne Software. Genomic DNA was amplified in triplicate using primers listed in Table 5, *GPR15* and *ZNF80* used as reference genes for analysis using ΔΔCT method[104].

For RT-qPCR reactions, RNA was extracted according to manufacturer's instructions using EZ-10 Total RNA Miniprep kit (Biobasic, BS88583). A one-step qPCR reaction was performed using Luna Universal One-Step RT-qPCR kit (New England Biolabs, E3005) on a CFX Connect Real Time qPCR system (Bio Rad). qPCR was performed using primers listed in Table 6, each sample in triplicate and data was analyzed using ΔΔCT method with *GAPHD* as an expression control.

Table 4. RT-qPCR primers

Gene	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Size (bp)
SMG1	TGGTGGAGAGTTACGCAGTC	AACTCTAAGGCTTTTACCTTTTTCA	174
Gapdh	GTCTCCTCTGACTTCAACAGCG	ACCACCCTGTTGCTGTAGCCAA	131

Table 5. qPCR copy number primers

Gene	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Size (bp)
SMG1	GGAAGGAATGCAACATCAGT	CACCACCAAAGATGGCACAA	84
SMG1	TTGGACAGCCTGGGTGGGAA	GTGAACAGCTCCTCGCCCTT	111
GPR15	GGTCCCTGGTGGCCTTAATT	TTGCTGGTAATGGGCACACA	101
ZNF80	CTGTGACCTGCAGCTCATCCT	TAAGTTCTCTGACGTTGACTGATGTG	120

Table 6. MiSeq Primers

Gene	Forward Primer (5' → 3')	Reverse Primer (5' → 3')	Size (bp)
SMG1	ACACTGACGACAT GGTTCTACAGTCTG CTGATTGTAATTCAGCAG	TACGGTAGCAGAGACT TGGTCTCACCCAGGCTGTCCAACCTT	194

Barcoding adaptor

siRNA transfection

Cells were seeded in 6-well or 24-well plates (7 x 10^5 and 3 x 10^5 cells per well, respectively), in DMEM lacking antibiotic-antimycotic solution. After 24 hours, cells were transfected with siRNA duplexes (Origene, SR307966) either the scramble or siRNA C at a final concentration of 10μM with lipofectamine RNAiMax (Thermofisher, 13778). After 48 hours knockdown was quantified using Guava flow cytometer (Luminex).

Transwell migration assay

To measure cell migration, VWR 24-well Polyester Membrane transwell plate inserts with 8.0um pore sizes (VWR 10769-242) were used as previously described[105]. Five hundred thousand cells were seeded in complete DMEM in a 6-well plate, after 24 hours medium was removed, cells were washed with 1X PBS and serum-free DMEM was added. After 24 hours of serum-starvation, cells were trypsinized in 0.25X 0.05% trypsin-edta, and resuspended in DMEM supplemented with 1% FBS. In each well, 600μL of 10% FBS-containing DMEM was added below the transwell insert and 1 x 10^5 cells in 100μL of 1% FBS DMEM was added to the transwell insert. The assay was incubated for either 48 or 72 hours. After incubation, transwells were swabbed, cells were fixed with 70% ethanol, and stained with 0.02% crystal violet (Sigma Aldrich, C6158). Stained transwells were imaged and cells that had migrated to the bottom of the transwell were quantified using imageJ or HALO software (Indica Labs). Experiment was performed in triplicate.

Cell proliferation and wound healing assays

Approximately 1 x 10^4 cells were seeded into 24 well flat bottom plates and proliferation of cells was detected by the Incucyte S3 Live Cell Analysis System (Satorius). Cells were grown for 72 and images were taken every 6 hours. Proliferation and doubling time (DT) were assessed by cell confluency and calculated with the formula: $Hours * Log(2) \div Log(Final\ Confluency) - Log(Initial\ Confluency)$.

To assess wound healing, 8.5 x 10^5 cells were seeded in each well of a 6-well plate such that cells were 90% confluent after 24 hours. A scratch was introduced using a

pipette tip in the middle of the well and wells were imaged under a microscope once a day at 24 hours intervals.

Clonogenic survival assay

Cells were seeded at 100, 200, 300, 400, and 800 cells per well in triplicate in a 6 well plate. Twenty four hours after seeding cells were irradiated at 2, 3, 4, and 5 gray, corresponding to plates seeded with 200, 300, 400 and 800 cells per well, respectively. Cells were cultured for 12 days and colonies that formed were fixed using 10% buffered formalin (Fisher Scientific SF100) and stained using 0.01% crystal violet (Fisher Scientific C581). Colonies of at least 50 cells were then counted manually and analyzed as described previously[106].

NMD reporter assay

Wild type or clonal 293T cell lines were seeded in a 6-well plate to reach 50-70% confluency at time of transfection ~750,000 cells/well. pKC-4.06 plates to reach ~75% confluency at the time of transfection. pKC-4.06 (FLuc-ß-globin-PTCstop) or pKC-4.04 (FLuc-ß-globin control) were co-transfected with pcDNA3-HCV-hRL using PEI (Figure 2A). PEI was prepared as previously described[107] to a stock concentration of 5mg/mL, and 5μg(1μL) of PEI was used per 1μg of DNA, 1μg of either experimental vector was co-transfected with 0.5μg HCV-hRL diluted in OptiMEM (Thermofisher 319850). Twenty four hours post-transfection, cell were reseeded in 24-well plates and 24 hours later were treated with 7.5μM of compound 52a, or lysed with passive lysis buffer (Biotium 99912) and luciferase activity was measured with luminometer (Promega). Analysis was conducted according to formulas in Figure 2B.

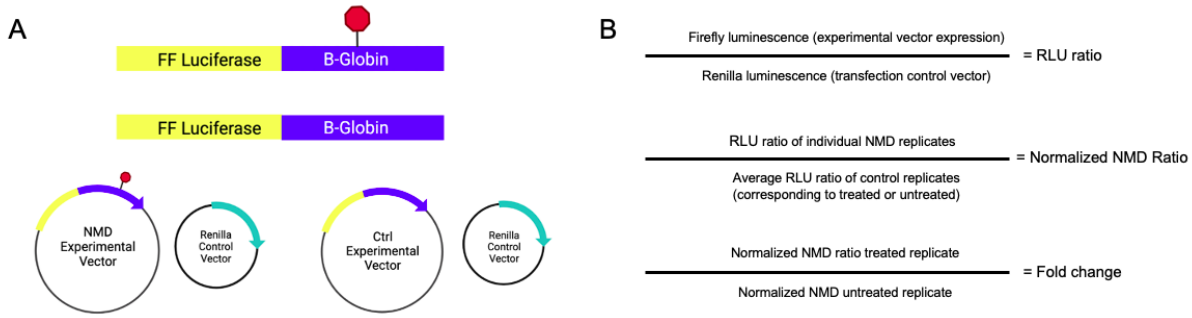

Figure 2. Nonsense mediated decay assay vectors and analysis. A) Schematic of vectors of NMD assay, firefly luciferase fused to beta globin gene with (NMD experimental vector pKC-4.06) or without (control experimental vector pKC-4.04) a premature termination codon. Firefly luciferase experimental vectors were co-transfected with a renilla transfection control vector (pcDNA3-HCV-hRL). B) Step-by-step analysis of NMD assay and normalization strategy.

In vivo tail vein metastasis assay

Cell were seeded 24 hours in advance such that they would be at 50-70% confluency at the time of harvesting. Cells were trypsinized 24 hours later, resuspended in complete DMEM, spun and resuspended in 1X PBS and 0.5μM EDTA (Thermofisher scientific 15575020) and treated for 5 minutes. Cells were spun again and resuspended in 1X HBSS (Wisent Bioproducts 311153), then passed through a 70μM cell filter. Cells were spun again and resuspended in 1X HBSS, counted and diluted to a concentration of 3 x 10^6 cells/mL. SCID beige mice were injected with 100μL of the cell suspension for a total of 3 x 10^5 cells injected. Mice were weighed once a week and sacrificed and necropsied when they reached humane endpoint. During necropsy the lungs, liver, and any masses were removed and fixed in 10% buffered formalin (Fisher Scientific SF100), and after 24-48 hours formalin was changed to 70% ethanol. Organs and masses were embedded in paraffin and H&E stained by the Histology Core Facility at the Rosalind and Morris Goodman Cancer Research Centre. Animal studies were approved by the McGill University Animal Care Committee, and conducted in accordance with animal research guidelines.

Chapter III: Results

Screening and attempted knock-in of c.4249A>G mutation into Hap1 Cell Lines

U2OS and HAP1 cells were co-transfected with a DNA donor template plasmid and a plasmid expressing spCas9, target sgRNA, and GFP. Forty-eight hours post-transfection, cells were single cell sorted by FACS for GFP expression, which would indicate successful transfection, and were cultured in 96 well plates until they could be expanded to 24-well plates for DNA extraction. DNA extraction followed by PCR amplification of the region of interest and Sanger sequencing was performed on 188 HAP1 clones, while few U2OS clones survived to this stage. The U2OS cell line is derived from osteosarcoma and was prioritized for the knock-in cell lines as previous studies on *SMG1* have been carried out in U2OS cells[108]. Due to the low survival of U2OS clones, we decided to attempt knock-in in HAP1 cells. The HAP1 cell line has human leukemia origin, derived from the KBM7 cell line, and it is nearly haploid. The nearly haploid nature of the HAP1 cell line makes it popular for CRISPR editing, as only one allele needs to be modified[109]. None of the sequenced clones had the mutation of interest, but there were alterations in 21.2% of clones sequenced, the most common alteration was in-frame deletions, specifically 3bp deletions (Figure 3).

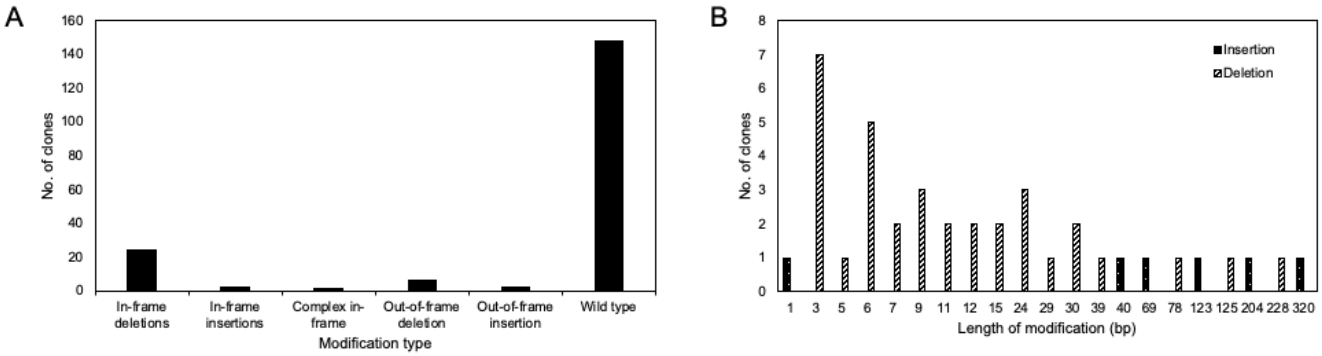

Figure 3. Screening sequencing results of single clones transfected with *SMG1* c.4249A>G CRISPR constructs. 188 single clones were screened using sanger sequencing to detect knock in of variant c.4249A>G. A) Types of modifications detected in screened clones. The majority of cell screened were wild-type and the most common modification was in-frame deletion. B) Frequency of the length and type of modification detected.

Generation of cell lines harbouring c.10921A>G mutations

HAP1, U2OS, and HEK293T cells were co-transfected with a DNA template plasmid and a plasmid expressing Cas9 and target sgRNA. HEK293T cells are immortalized human embryonic kidney cells and were selected for knock-in generation due to their hardy nature and ease of transfection. Cells were cultured for one week after transfection and then sorted for GFP expression, which would indicate successful knock-in of the mutation construct. One week after transfection, no GFP-positive cells were detected in HAP1 cells lines, and 0.2% and 0.03% were found in HEK293T and U2OS cells, respectively. Five U2OS clones that were passaged and expanded up to a 24-well plate were visualized to be GFP-positive. Of the 5 GFP+ clones, none were found to have the mutation inserted in the correct genomic location. Of the HEK293T cells single-sorted and grown into cell lines, clone A3 had a mixed population of GFP positive and negative cells, and these cells were again sorted into single cells to generate single GFP+ clones. This gave rise to clones A6, A9, and A17, all of which were confirmed to have the mutation of interest in the correction genomic locus through Sanger sequencing (Figure 4a). To determine whether the clones were heterozygous or homozygous, primers outside the donor template region were used to amplify and discriminate between alleles by size, the mutant allele with a GFP insertion was expected to yield a product of approximately 2000bp, whereas the wild type allele was expected to yield a product of 1600bp (Figure 4B). Clones A6, A9, and A17 had amplicons of both 2000bp and 1600bp, whereas the WT cell line did not have a 2000bp amplicon, indicating the clones are heterozygous.

Clones A.6, A.9, and A.17 were selected to go forward with functional characterization and were further characterized genomically. All clones were sequenced not only around the mutation of interest, but also around the DNA template homology arms to ensure no unintended mutations were present. To confirm no more than two copies of *SMG1* gene were present in the clone genomes, qPCR of genomic DNA was performed and clones were compared to each other and wild type HEK293T cells and qPCR amplification was normalized to known diploid cell line MRC-5. Average relative expression of *SMG1* amplicon was between 0.623 number and 0.690 number for all *SMG1* clones normalized to *GPR15* (Figure 4C) and between 0.640 and 0.890 when normalized to ZNF80, indicating there is likely no more than 2 copies of *SMG1* (Figure 4D). RT-qPCR showed that all clones expressed *SMG1* on the mRNA level (Figure 4E).

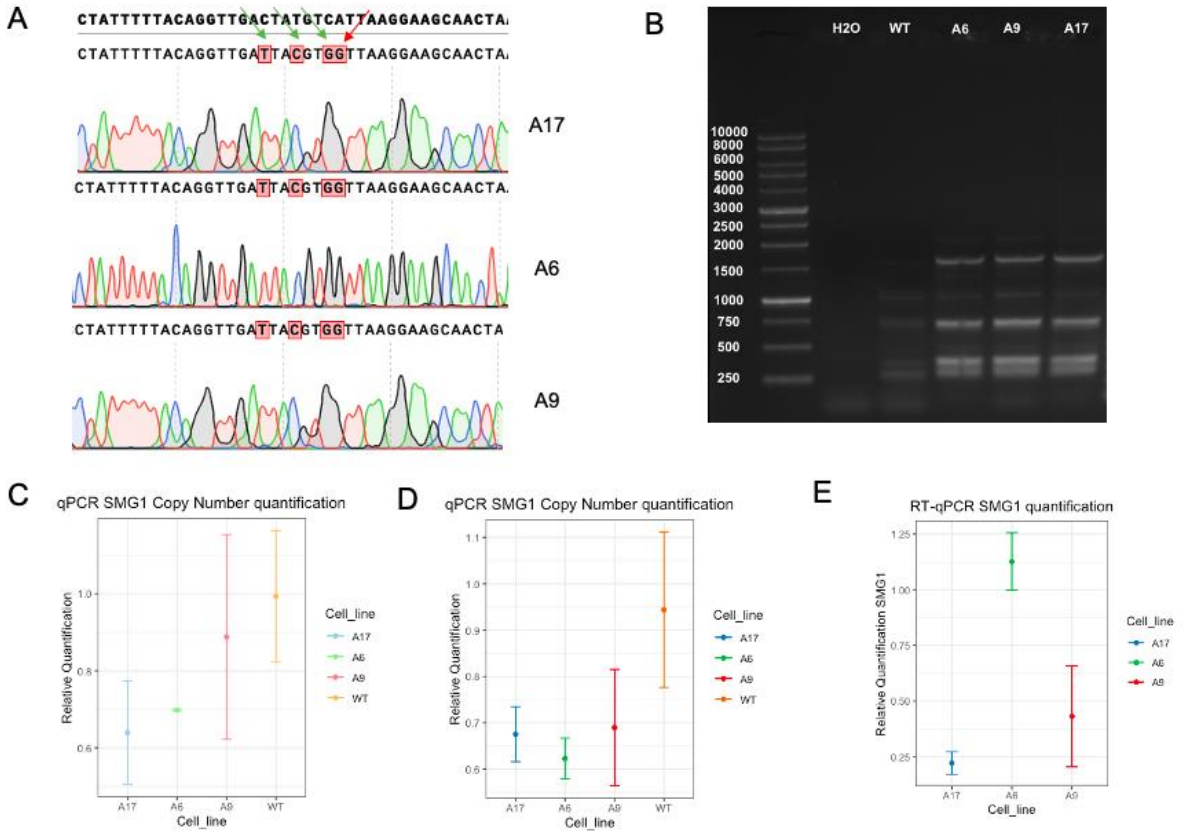

Figure 4. Genomic characterization of c.10921A>G clone cell lines. A) Sanger sequencing of area surrounding c.10921A>G mutation, green arrows pointing to silent mutations in the protospacer region, red arrow pointing to missense mutation. B) PCR amplification of area outside homology arms showing clones are heterozygous, with expected bands at approximately 1600bp for c.10921A>G allele with GFP insert and 725 for WT allele. C) Relative quantification of copy number by qPCR, relative to diploid cell line MRC-5 using *GPR15* as a reference gene. D) Relative quantification of copy number by qPCR, relative to diploid cell line MRC-5 using *ZNF80* as a reference gene. E) Relative quantification of *SMG1* expression by RT-qPCR, expression relative to WT HEK293T cell line.

Characterization of doubling time and wound-healing

Doubling time was calculated from confluency Incucyte measurements. Mean doubling time from fastest time point, the 48-72 hour time-points were calculated as 25.12h, 26.98h, 18.44h, and 18.804 corresponding to cell line A6, A9, A17, and wild type,

respectively. Doubling time was significantly higher in all c.10921A>G cell lines in comparison to wild type (WT) at 48 hours, (A6 p=.025, A9 p =7.12 x 10^{-6}, A17 p =0.043 , while only cell line A9 had increased doubling time at the 72 hour time point (p=0.033) (Figure 5A).

In order to evaluate doubling time in cells with reduced SMG1 expression and therefor function, siRNAs either against *SMG1* or scramble siRNAs were transfected. SMG1 knockdown was quantified by flow cytometry and at least 50% reduction of GFP expression was observed after 48 hours in all clones. The average doubling time of cells transfected with *SMG1* or scramble at their fastest respective DTs were 28.48h and 26.23h in A6, 24.84h and 24.26h in A9, 26.417h and 23.13h in A17, and 15.64h and 17.29h in WT (Figure 5B). Addition of siRNAs against *SMG1* significantly decreased DT in WT (p<0.01) at the 24h time point, but did not have a significant effect on clones A6, A9, A17 at any time point, or at any other time point in the WT cell line.

Wound healing was characterized by a scratch assay and microscope images were taken at 24 hour intervals over 72 hours (Figure 6). All wounds had closed by 72 hours.

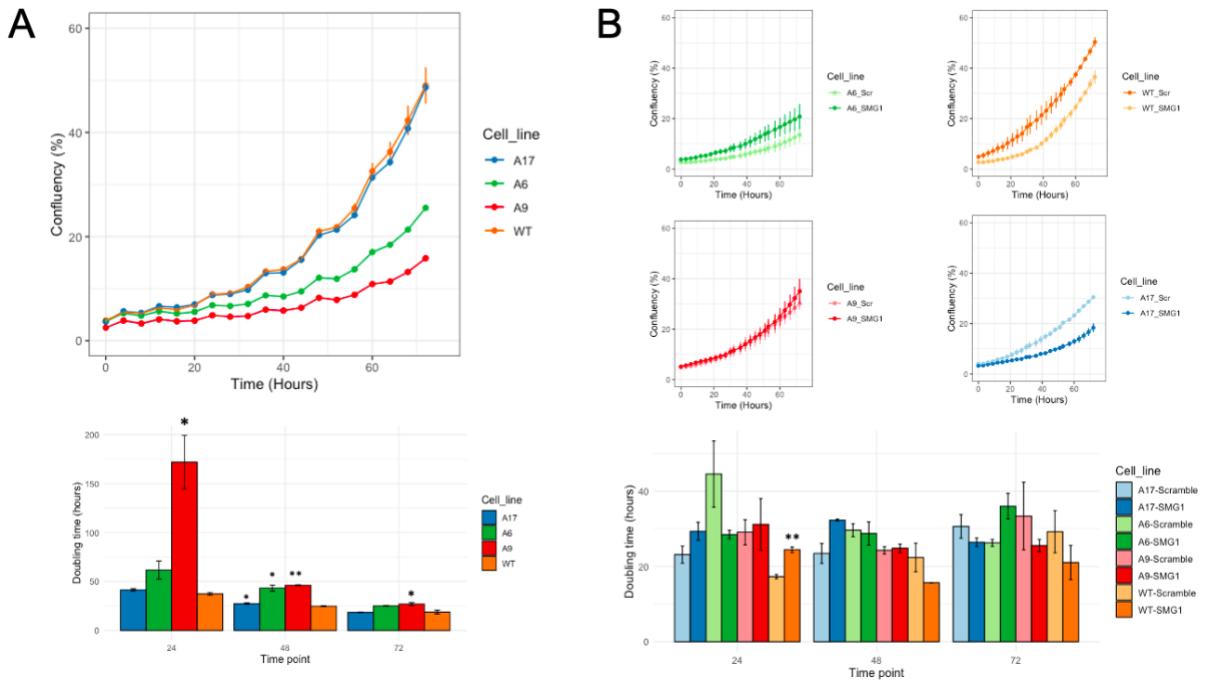

Figure 5. *SMG1* c.10921A>G clone proliferation and doubling times. A) Proliferation measured by confluency over 72 hours for c.10921A>G cell lines A6, A9, A17 and wild type

HEK293T cell line and corresponding doubling time. Doubling time was calculated in 24-hour intervals from 0-24 hour (24 hour time point), 24-48 hours (48 hour time point), and 48-72 hours (72 hour time point). Doubling times were compared using unpaired students two-tailed t-test, *P<0.05. B) Proliferation and doubling time after the transfection of scramble or *SMG1*-targeting siRNAs.

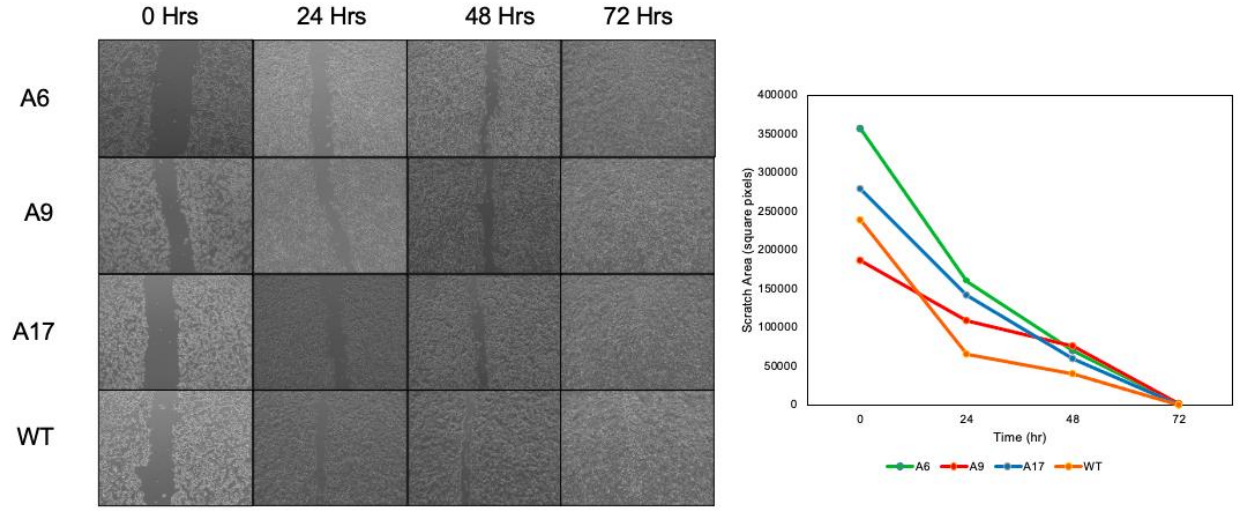

Figure 6. Wound healing measured by scratch assay in HEK293T wild type and c.10921A>G *SMG1* clones. A) Wound healing imaged in 24-hour intervals for 72 hours following a scratch introduced by a p200 pipette tip. B) Quantification of wound area by ImageJ analysis.

Characterization of migration

Migration was evaluated using the transwell migration assay, cells were starved of serum and seeded in triplicate in the upper chamber of the transwell with serum-rich medium in the lower chamber. Migration was quantified manually using ImageJ during initial experiments, which showed increased migration in all clone cell lines in comparison to wild type after 72 hours (Figure 7B) but migration was only significantly higher in clones A6 and A17 after 48 hours in comparison to wild type (Figure 7A).

The addition of siRNAs against *SMG1* increased migration significantly in the WT cell line only. In siRNA migration experiments, migration was evaluated using HALO software in an effort to improve accuracy, the software is trained to recognize the area

stained by crystal violet and the percentage image stained is reported, rather than cell number manually counted with ImageJ. The mean percentage area covered by migrated cells is 66.62% and 91.48% in cell line A6, 39.76% and 22.56% cell line A9, 72.75% and 74.30% in cell line A17, and 37.16% and 16.63% in WT cell line, transfected *SMG1* or scramble siRNA, respectively after 48 hours of incubation (Figure 7A). The addition of the siRNA against *SMG1* significantly increased migration in comparison to cells transfected with scramble siRNA after 48 hours of incubation in clonal cell line A9 (p=0.021) and WT (p=0.003), but did not have significant effect on cell lines A6 or A17. At 72 hours mean area with migrated cells was 81.41% in *SMG1*-siRNA transfected cells and 86.13% in scramble siRNA transfected cells in A6, 70.5% and 80.2%in A9, 91.80% and 85.50% in A17, and 83.80% and 83.74% in WT (Figure 7B). There was a significant increase in migration in SMG1-siRNA transfected cells compared to scramble-siRNA transfected cells only in the A17 clonal cell line (p=0.017) after 72 hours of incubation.

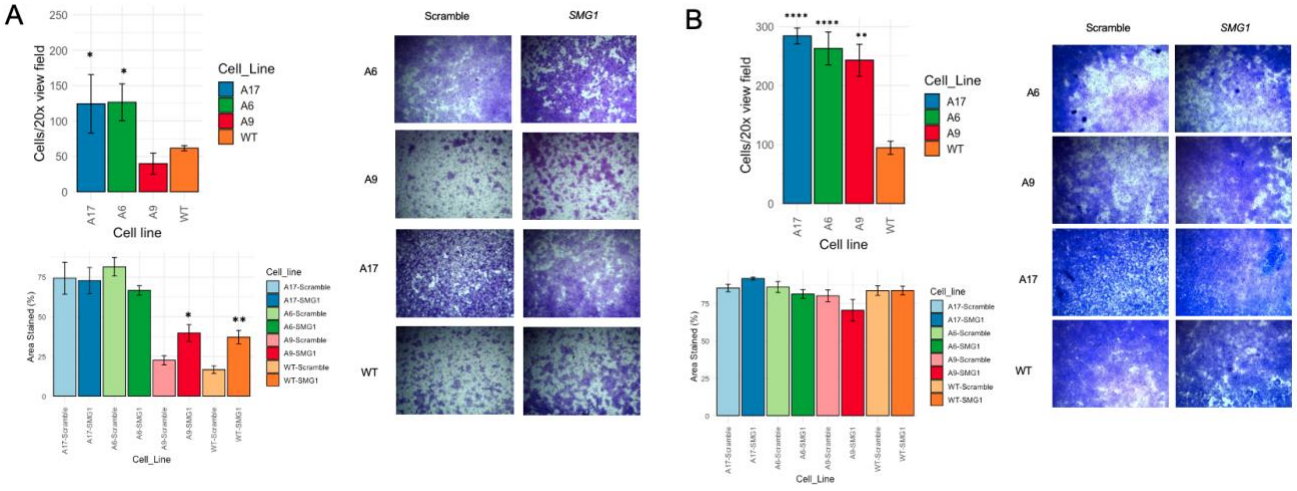

Figure 7. Trans-well migration of clonal cell lines. Transwell migration with and without the addition of scramble or SMG1-targeting siRNAs after 48 (A) and 72(B) hours.

Survival after exposure to IR

Survival after exposure to ionizing radiation was evaluated in cell lines using a clonogenic survival assay. Cells were seeded in triplicate an increasing number corresponding to dose of IR. Initial experiments using 3, 5, and 7 Gy showed a significant decrease in survival of cell lines A6 and A9 in comparison to WT ($p<0.05$), plotted using

weighted, stratified quadratic regression (Figure 8A). The survival curves suggested the effect may be most obvious at lower dosages of radiation and there were small surviving fractions at 7 Gy across cell lines, dosages were adjusted to 2, 3, 4, and 5 Gy of IR for repeat experiments. At adjusted dosages, significantly reduced survival was observed in both A9 (p=0.0003) and A17 (0.0234) clones in comparison to WT, while A6 was not significantly different (p=0.0728) (Figure 8B).

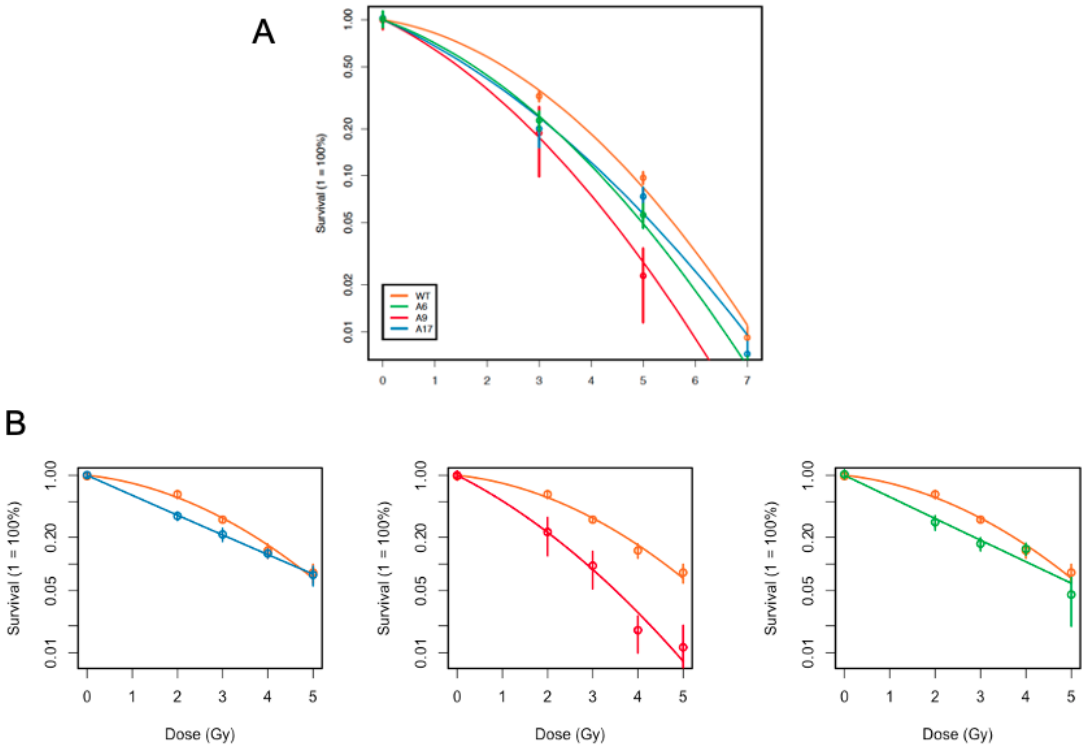

Figure 8. Survival after exposure to ionizing radiation. A) Survival curve of 3 c.10921A>G clones and WT cell line after exposure to 3, 5, and 7 Gy. Survival was fitted by a weighted stratified regression according to a linear-quadratic formula and p-values were calculated using ANOVA. B) Survival of individual clones compared to WT after 4 doses of ionizing radiation: 2, 3, 4, and 5 Gy.

Nonsense-mediated decay

Nonsense-mediated decay function was quantified using a dual luciferase system. Cells were co-transfected with renilla expression vector and a firefly luciferase vector fused to either an in-frame beta-globin gene, or a beta-globin gene harbouring a

premature stop codon (Figure 2A). Cells were transfected and then seeded into 6 technical replicates, and firefly luminescence was normalized to renilla luminescence to control for transfection efficiency and number of cells (Figure 2B). The luminescence ratios from cells transfected with the NMD vector were normalized to the cells transfected with the control vector to give an NMD expression ratio for each cell line. In 3 experiments, A17 was found to have a significantly higher normalized NMD ratio than WT ($p>0.05$), and in 2 experiments clones A6 and A9 had significantly higher normalized NMD ratios as well ($p>0.05$) (Figure 9).

To determine whether the presence of a wild type allele could be further masking the effect of the mutant *SMG1* allele on NMD, NMD inhibitor 52A was added to the assay. Twenty four hours after transfection of NMD vectors, cells were seeded into 24-well plates, and 24 hours after that compound 52a was added at 7.5uM concentration and incubated for 6 hours before lysing and measuring luminescence. Across experiments, wild type cells had higher expression from the NMD vector following addition of 52a, indicating increased inhibition by 52a (Figure 10A).

To determine the effect of reduced *SMG1* expression, and therefor activity, on NMD in both clones and wild type cell lines, cells were transfected with either a scramble siRNA or an siRNA against *SMG1*, after 48 hours they were transfected with NMD expression vectors, then split into 6 technical replicates and 24 hours later they were lysed and luminescence was read. The addition of siRNA against *SMG1* significantly increased expression from NMD vector in all cell lines, A6 ($p=0.0002$), A9 ($p=0.0069$), A17 ($p=0.0058$), and wild type ($p=0.00560$). However, there was no significant difference in the fold change between scramble and *SMG1* siRNA, NMD vector transfected cells, between wild type and clone cell lines (Figure 10B).

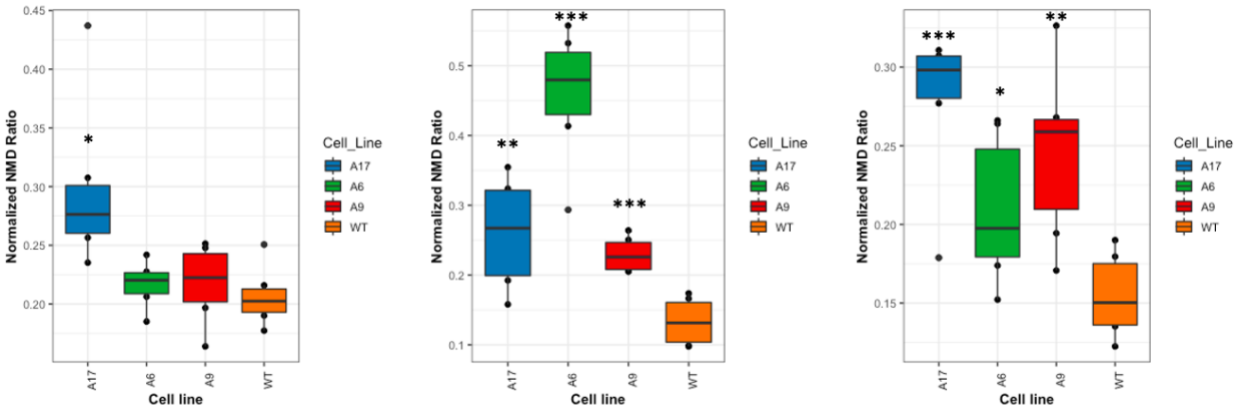

Figure 9. NMD function of c.10921A>G *SMG1* clones measured by luciferase assay. RLU ratio calculated by RLU firefly luciferase expressed by the experimental vector (either NMD or control) divided by renilla luciferase expressed by the transfection control vector. RLU ratio of NMD vector-transfected samples were normalized to the RLU ratio of experimental control vector-transfected samples and plotted. Each plot represents one experiment and significance was determined using students t-test (*p<0.05, **p<0.01, ***p<0.001).

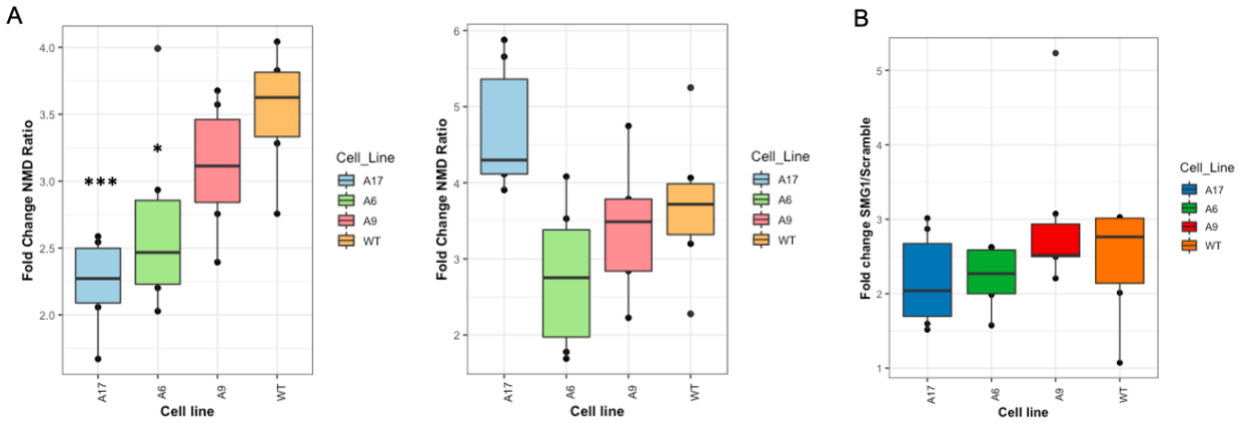

Figure 10. Inhibition of NMD by 52A or *SMG1* siRNAs across 3 experiments. Fold change of normalized NMD expression ratio between treated and untreated cells with 52a compound or *SMG1* siRNA. RLU ratio calculated by RLU firefly luciferase expressed by the experimental vector (either NMD or control) divided by renilla luciferase expressed by the transfection control vector. RLU ratio of NMD vector-transfected samples were normalized to the treatment corresponding

RLU ratio of experimental control vector-transfected samples. Normalized NMD ratio of 52a-treated over untreated cells was used to calculate the fold change NMD ratio. A) Clone and HEK293T WT response to 52a inhibitor across two experiments. B) Effect of *SMG1* siRNA on NMD as measured by fold change between *SMG1*-siRNA and Scramble-siRNA NMD RLU.

Development of masses from tail vein injection

As an extension to the *in vitro* migration assay, an *in vivo* experimental metastasis assay was utilized to determine whether c.10921A>G SMG1-mutated cell lines would form tumours at an increased rate or form more tumours than the WT cell line. Male adult beige SCID mice were injected with either 3 x 10^5 cells of A9 (n=3), A17 (n=3), or WT (n=3) cell line via tail vein. At week 15, 6 of the 9 mice in this pilot experiment have reached endpoint (Table 7). Two mice from each of the 3 treatment groups have reached endpoint, and 5 of the 6 mice that had reached endpoint developed mass lesions, most frequently these lesions developed in the abdomen. The remaining 3 mice have not developed tumours at the time of the submission of this thesis and will continue to be monitored for any mass development.

Table 7. Development of tumours *in vivo* metastasis assay

Mouse	Cell Line Injected	Survival (weeks post-injection)	Mass Lesions
A	A9	8	Neck and abdomen
B	A9	11	Abdomen, lungs, thymus
D	A17	6	No
E	A17	12	Abdomen
G	WT	15	Neck
I	WT	7	Abdomen

Chapter IV: Discussion

PDAC is one of the deadliest solid malignancies with a devastatingly low 5-year survival of less than 10%[110]. However, there is growing optimism that the stepwise progression modality leading to precursor lesions of PDAC will allow for early detection and intervention in individuals at risk[111]. Up to 10% of PDAC cases cluster in families, and the majority of these cases do not have a known genetic cause[32]. Thus, elucidating the genetic basis for the unaccounted fraction of FPC will enable expansion of gene-based surveillance strategies for detection of early stage PDACs and their precursor lesions in at-risk kindred.

The interest in genetic predisposition of PDAC has led to several major breakthroughs in PDAC genetics, including exome-wide association studies identifying variants in DNA repair genes associated with PDAC, targeted sequencing panels showing 13.5% of PDAC patients harbour germline pathogenic mutations, and prospective kindred studies concluding a missing major PDAC susceptibility/early age of onset gene[33,112–114]. More recently, our lab performed a region-based gene association study which identified *SMG1* as a novel PDAC predisposition gene[40]. *SMG1* has been associated with studies showing its downregulation either by promoter hypermethylation or noncoding RNAs in most cases leading to a more aggressive tumour[48–50,52]. The only published work on *SMG1* and predisposition to cancer used a mouse model and concluded *Smg1* was a haploinsufficient TSG, showing the knockout of one copy of *Smg1* was sufficient to predispose to the development of tumours[115]. The exact mechanism by which *SMG1* can lead to predisposition or association with cancer is unclear. Further, there have been no studies on germline *SMG1* mutations and cancer, and whether the missense *SMG1* variants associated with PDAC identified in Wong et al. have functional consequences is unknown.

The advent of CRISPR gene editing has allowed for relatively efficient generation of *in vitro* models for variants identified in databases such as ClinVar or in clinical studies. While Wong et al. identified scored pathogenicity of identified variants using *in silico* predictions, functional validation is needed before pursuing *SMG1* in a clinical context. To functionally validate select mutations identified in Wong et al., we used CRISPR/Cas9 gene editing to generate cell lines harbouring the c.10921A>G mutation and attempted to generate cell lines harbouring the c.4249A>G mutation.

Our results show that there are functional differences between one of the identified *SMG1* variants, variant c.10921A>G, and wild type *SMG1* cells. The results of the copy number qPCR and genotyping PCR indicate that our clones have two copies of the *SMG1* gene: presumably, one WT allele and one c.10921A>G allele. However, neither the qPCR for copy number nor the RT-qPCR for *SMG1* expression can discriminate between WT and mutated alleles, as qPCR amplicons are limited to approximately 200bp. This is too short to place a primer complementary to a sequence unique to the mutant allele and the second primer outside the DNA template homology arms. Thus, the qPCR for mutant alleles would not be able to discriminate between a mutant allele at the correct locus and the DNA template integrated at incorrect loci. The qPCR for copy number and the RT-qPCR for expression are a product of the cumulative *SMG1* copy number or expression, regardless of mutation status. The WT allele copy number can be estimated by quantifying the copy number of the clonal and WT cell lines relative to known diploid cell line, MRC-5. In order to quantify the copy number of the *SMG1* mutant allele, a control of known mutant allele copy number is needed. However, the results of the genotyping PCR and the copy number qPCR together indicate our clones are likely heterozygous, as the copy number qPCR indicate our clones had no more than 2 alleles of *SMG1* and the genotyping PCR showed the presence of both a WT and the c.10921A>G allele. A northern blot with probes specific to both the wild type and mutant allele mRNA could best inform on the expression of *SMG1* in the clonal cell lines, and could explain some of the clonal differences. In a heterozygous mutant cell line, a change in function is encouraging of haploinsufficiency, as it indicates that only one allele needs to be mutated in order to alter function. Our clonal cell lines bearing the c.10921A>G mutations demonstrated increased migration, increased doubling time (decreased proliferation), decreased survival after exposure to IR, and reduced NMD function and reduced sensitivity to NMD inhibition by the compound 52a. Taken together these results support the hypothesis that the c.10921A>G missense variant reduces *SMG1* function and further asserts *SMG1* as a haploinsufficient TSG, as reduced function of *SMG1* is the result of a heterozygous mutation.

As a presumed tumour suppressor, we expected *SMG1* to harbour loss-of-function mutations, which increase both migration and proliferation. A study on miRNA-18a in nasopharyngeal cancer found that suppression of *SMG1* increased colony formation as well as invasion[52]. Similarly, in gastric carcinoma, *SMG1* suppression was found to increase invasion, proliferation, and motility[47]. A limitation of the current clinical studies investigating *SMG1* is the focus on miRNAs, and by extension, experimental design that indirectly assess *SMG1* downregulation rather than targeting *SMG1* directly. This renders the comparison between our study and others difficult. Generally, it is expected that both migration and proliferation would increase follow the suppression or loss-of-function of a presumed TSG. Interestingly, that is not what we observed. Migration was significantly higher in two of three clone cell lines across several experiments, while doubling time was increased in all clones at the 48 hour time point, indicating a decrease in proliferation. Three clonal cell lines seems typical of CRISPR/Cas9 studies in order to control for clonal variability, though the exact number of clonal cell lines is not always reported. Moreover, this number may be even higher for CRISPR knock-out experiments, but the difficulty of generating a knock-in cell line could limit the number of clones used. In the migration, NMD and radiosensitivity experiments, two of the three clones differed significantly from the WT cell line, and while having all clones differ from the WT cell line would increase our confidence in these results, two of three clones in agreement is reassuring. The variation between clones could be technical or due to difference in expression of *SMG1*, the heterozygosity of the clones, or genetic variation between clones. Due to the size of the SMG1 protein, western blotting is technically challenging, but we do intend to quantify SMG1 protein expression in the future to further characterize our clones. Expression of SMG1 protein is especially important to consider in heterozygous clones, as increased expression of the WT allele could mask the effects of the mutated allele and contribute to variation. One study which generated 15 mutant cell lines, 4 of which with targeted CRIPSR mutations, found that there was a high level of variability between the clonal cell lines, but the responses of individual cell lines were consistent across experiments[116]. Further, the authors of this study explored the genetic variation between clones and mutations accumulated between clonal generation and functional studies and found many

unique variants when comparing clones[116]. Investigating genetic variation between our clonal cell lines could explain some of the variation in our results.

Increased radiosensitivity in cells with reduced *SMG1* levels is a known phenomenon. The first study to characterize the function of *SMG1* in genotoxic stress response showed that inhibition of *SMG1* using siRNAs decreased survival after exposure to IR, likely due to its involvement with p53 activation[41]. Another study that showed decreased expression of *SMG1* in HPV-positive HNSCC demonstrated that the knockdown of *SMG1* in HPV-positive HNSCC cell lines decreased survival after exposure to IR due to increased induction of apoptosis[50]. Similarly, the increased radiosensitivity in our *SMG1* c.10921A>G cell lines is in line with what is expected with loss-of-function *SMG1* mutations. Further, the repetition of this experiment in the future with siRNAs against SMG1 will increase our confidence that the reduction of SMG1 function leads to an increase in radiosensitivity.

The reduction in NMD activity in the *SMG1*-mutated cell lines was unexpected. A study by Roberts et al. using a *Smg1*$^{gt/wt}$ mouse model with only one functioning *SMG1* allele showed no difference in NMD capabilities compared to *Smg1*$^{wt/wt}$ mouse. Given our cell line is heterozygous for c.10921A>G *SMG1* mutations, it was interesting to observe impaired NMD. This difference could be due to differences in experimental design; Roberts et al. used the endogenous target of NMD *Gas5*, while we used an exogenous target[115].

Previous *in vitro* studies have shown reduction of *SMG1* by siRNAs reduces NMD function, however, these studies achieved greater than 50% knockdown[117]. The lessened effect of the 52a compound on *SMG1* c.10921A>G cell lines is in line with the observed decrease in NMD function: if NMD is reduced at baseline, then a compound targeting upstream of SMG1 function is expected to have less of an effect.

The addition of siRNAs targeting *SMG1* in WT cell lines support the observed differences in migration, proliferation, and NMD assays in c.10921A>G mutant *SMG1* cell lines. Though the mode of action of loss-of-function mutations is not the same as siRNAs, the end result is the same: a decrease in functional protein levels. After the transfection of *SMG1* siRNAs in WT cells and therefore the decrease in functional SMG1 protein levels, doubling time increased, similar to the increased doubling time of *SMG1*

c.10921A>G mutant cell lines at baseline. Parental cell lines followed the same trends relative to each other in the siRNA experiments as in the untreated experiments, though the doubling times were longer in the untreated experiments. Similarly, the significant increase in migration in WT cells transfected with *SMG*1 siRNA supports the observed increased migration in mutant cell lines, and further supports loss-of-function in the c.10921A>G mutation. While we did not observe an additive effect of the *SMG1* siRNAs in the clonal cell lines, as there was no difference between scramble and *SMG1* siRNA treated cells except in the case of clone A9 and increased migration, which could be attributed to the heterozygosity of the clones. In the case of heterozygosity, siRNAs target both the WT and mutated allele; if the siRNAs only targeted the WT allele, we would likely see an additive effect, assuming c.10921A>G truly is a loss-of-function variant.

Our inability to generate a cell line harbouring the *SMG1* variant c.4249A>G merits discussion. *SMG1* has been determined by project DRIVE to be essential in cell lines. Project DRIVE used RNAi to determine viability effects on mRNA knockdown of ~8,000 genes and used an average of 20 shRNAs per gene[102]. The RSA function of project DRIVE uses data from all shRNA experiments to determine statistically whether a gene is essential to cell survival. Genes with RSA with less than -3 in more than 50% of cell lines was determined to be essential[102]. *SMG1* has an RSA of less than -3 in 57% (218 of 382) cell lines tested, indicating it is essential for viability in cell lines[102]. Thus, the difficulty in growing U2OS clones, and our inability to generate a c.4249A>G *SMG1* knock-in in HAP1 cells may be due to generation of indels creating stop codons and knocking out *SMG1*. We demonstrated reasonable efficiency by the targeting sgRNA and Cas9, observing modification in approximately 1 in 5 clones screened, with most modifications in the screened clones being short in-frame deletions. The HAP1 cell line, which is the cell line used to attempt to generate c.4249A>G clones, is a modified KBM-7 cell line that is almost completely haploid, and therefore frameshift mutations resulting in a nonsense mutation in an essential gene should be lethal. However HAP1 cells will duplicate their genome at high passage. It is possible that the clones with frameshift mutations had duplicated their genome by the time they were screened, as was the case with a clone with an 8bp deletion that investigated after sequencing. It is also possible that a missense mutation that knocks out the function of an essential gene would be lethal

in a haploid cell line. The inability to generate a *SMG1* c.4249A>G clone may therefore indicate the toxic effects of this missense mutation, or it may just illustrate the difficulty of generating CRISPR knock-ins.

Chapter V: Conclusions and Future Directions

Taken together, the results presented support our hypothesis that the missense *SMG1* variant c.10921A>G does have functional consequences. However, it is difficult to resolve this contribution to PDAC predisposition. The combination of increased migration with decreased proliferation in our c.10921A>G *SMG1* cell lines could be suggestive of a contribution in the progression of tumourigenesis rather than establishment of PDAC. This may fit within with the progressive model of PDAC pathogenesis whereby mutations accumulate in a step-wise manner. This could also suggest c.10921A>G as a low penetrance allele, perhaps contributing to PDAC establishment in some individuals and not others, or a tissue-specific observation, as *SMG1* loss-of-function has been shown to increase both migration and proliferation in other cell types. Further, our results indicate that *SMG1* c.10921A>G mutations could be advantageous in metastatic progression. Both increased proliferation and migration are advantageous in cancer, but at certain time points in tumourigenesis they are most advantageous separately. A study of the YB-1 transcription factor showed that it contributes to tumourigenesis through coordinated repression of proliferative factors while upregulating mesenchymal factors to facilitate epithelial-mesenchymal transition (EMT)[118]. The authors then went on to propose that during EMT cancer cells must postpone growth and instead upregulate factors to cope with the stress of metastasis[119]. This would be in line with our results.

While a cellular model is useful for investigating functional consequences of missense mutations, the *Smg1* gene trap mouse model developed by Roberts et al. demonstrates how an *in vivo* model may be important to further this research and investigate how *SMG1* missense mutations contribute to PDAC predisposition. Roberts and colleagues examined NMD function and stress response, but found no differences between *Smg1*$^{wt/wt}$ and *Smg1*$^{gt/wt}$; they instead detected low levels of inflammation in *Smg1*$^{gt/wt}$ mouse undetected in the WT mice[115]. Increased inflammation was attributed to cellular damage by the presence of reactive oxygen and nitrogen species[115]. Reduced SMG1 levels or SMG1 function predisposing to PDAC by way of inflammation is in keeping with general cancer development, but also with increased PDAC risk in individuals with diabetes or pancreatitis[120]. Inflammation and immune processes cannot be fully characterized in a cellular model, and although the generation of a GEMM harbouring *SMG1* missense mutations is labour intensive and may not be advised without

further functional evidence, techniques to deliver CRISPR constructs directly to the mouse pancreas have been described and could be informative on PDAC tumourigenesis in the background of a *KRAS*-mutated mouse[121]. Our *in vivo* studies were limited to a pilot metastasis experiment, which demonstrated both our c.10921A>G *SMG1* cell lines and wild type HEK293T cell lines can establish tumours *in vivo* via tail vein injection. HEK293T cells have been shown to be tumourigenic at high passage, but to our knowledge there have been no experimental metastasis models using HEK293T cells[122]. While there was no clear difference or pattern in the timing of tumour development or location of tumour development when comparing WT vs mutated cell lines, further analysis on the mass lesions formed during our experimental metastasis is underway. It is our hope that the histological characterization by immunohistochemistry will identify differences between tumours formed in mice injected with WT and mutated cell lines. Though there was no evidence of a decrease in NMD function in the *Smg1* gene trap mouse model, it would be informative to examine the effects of missense variants in *Smg1* on NMD *in vivo*. The study by Roberts et al. showed in certain tissues there was only a slight reduction in the full-length Smg1 protein, indicating there was some level of post-transcriptional regulation and that a missense variant resulting in reduced *Smg1* function could have a dominant negative effect. Indeed, exploring NMD *in vivo* is important since the role of NMD is multifaceted, especially in cancer. Normal NMD function works to prevent aberrant or truncated proteins from being produced, and has been found to have tumour suppressive qualities, regulating transcripts of growth, proliferation, differentiation, and survival factors[123]. Tumours can utilize NMD to exert selective pressures: negative selection against loss-of-function mutations in essential genes and oncogenes, but a positive selection for loss-of-function mutations in tumour suppressor genes[124]. One pan-cancer study found that mutations that elicit NMD target tumour suppressor genes at higher frequency, concluding that some tumours depend on functioning NMD in order to eliminate the function of tumour suppressor genes[125]. NMD may also function in antitumour immunity. Inhibition of NMD by shRNAs against *Upf2* or *Smg1* in mouse xenografts has been shown to decrease tumour volume by triggering antitumour immunity through the production of neoantigens[126]. This finding was replicated in a separate *in silico* study that found removal of frameshift mutations captured by NMD led to tumour

mutational burden becoming predictive of survival in 4 of the 5 cohorts examined[127]. Further, this led authors to suggest NMD suppresses both immune reactivity in tumours and response to immunotherapy[127]. Frameshift indels that can elude NMD are rare, but they have been found to associate with antitumour immunity, likely through production of neoantigens[128]. There is mounting evidence asserting the importance of the immune microenvironment in cancer, and NMD is a recent area of interest in immuno-oncology; determining how variants in the NMD factors, especially germline variants, affect the tumour immune microenvironment is an exciting and unexplored avenue.

To characterize additional *SMG1* variants associated with PDAC, a high throughput assay is advantageous. To prioritize variants identified in Wong et al. for further characterization by CRISPR, a “knockdown-overexpress” model could be used. This technique uses stable shRNA or siRNA to knockdown the endogenous protein of interest, and transiently expresses an exogenous cDNA with site directed mutagenesis of variants of interest. This approach has been used to evaluate *PALB2* variants of unknown significance identified in case-control consortiums and ClinVar for functional consequences[129]. This approach has also been used to model hotspot mutations identified in the *KLF5* gene in colorectal cancer and found that missense mutations in *KLF5* increase protein stability and by disrupting interactions with regulatory proteins[130].

Another consideration of the use of the *in vitro* cell model is the cell type used; the original aim of this research was to create U2OS cell lines harbouring *SMG1* missense variants, as key *SMG1* functional studies had been conducted in U2OS cell lines. However, few U2OS clones survived after FACS sorting, and all clones with the c.10921A>G-GFP construct appeared to have the construct inserted in the incorrect locus[41,69]. Moreover, the contrast in our results compared to the literature in regards to proliferation may demonstrate tissue specific differences in functions of *SMG1*, underscoring the importance of using a cell line most closely modelling the disease of interest.

In conclusion, our *SMG1* c.10921A>G variant cell lines have displayed functional differences in comparison to their wild type counterpart, demonstrating increased migration, decreased proliferation, decreased survival after genotoxic stress and

decreased nonsense-mediated decay. Though more research is needed to validate *SMG1* as a PDAC predisposition gene, this dissertation is an encouraging first step.

Chapter VI: Bibliography

1. Debi, U. *et al.* Pancreatic trauma: A concise review. *World Journal of Gastroenterology* **19**, 9003–9011 (2013).
2. Leung, P. S. Overview of the pancreas. *Advances in Experimental Medicine and Biology* **690**, 3–12 (2010).
3. Zhou, Q. & Melton, D. A. Pancreas regeneration. *Nature* **557**, 351–358 (2018).
4. Leung, P. S. Physiology of the pancreas. *Adv. Exp. Med. Biol.* **690**, 13–27 (2010).
5. Distler, M., Aust, D., Weitz, J., Pilarsky, C. & Grützmann, R. Precursor lesions for sporadic pancreatic cancer: PanIN, IPMN, and MCN. *BioMed Research International* **2014**, (2014).
6. Fischer, C. G. & Wood, L. D. From somatic mutation to early detection: insights from molecular characterization of pancreatic cancer precursor lesions. *Journal of Pathology* **246**, 395–404 (2018).
7. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
8. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).
9. Notta, F. *et al.* A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. *Nature* **538**, 378–382 (2016).
10. Yao, W., Maitra, A. & Ying, H. Recent insights into the biology of pancreatic cancer. (2020). doi:10.1016/S1470-2045(20)30088-7
11. Brenner, D. R. *et al.* Projected estimates of cancer in Canada in 2020. *CMAJ* **192**, E199–E205 (2020).
12. Paulson, A. S., Tran Cao, H. S., Tempero, M. A. & Lowy, A. M. Therapeutic advances in pancreatic cancer. *Gastroenterology* **144**, 1316–1326 (2013).
13. Jones, R. P. *et al.* Patterns of Recurrence After Resection of Pancreatic Ductal Adenocarcinoma A Secondary Analysis of the ESPAC-4 Randomized Adjuvant Chemotherapy Trial Supplemental content. *JAMA Surg* **154**, 1038–1048 (2019).
14. van Veldhuisen, E. *et al.* Locally advanced pancreatic cancer: Work-up, staging, and local intervention strategies. *Cancers* **11**, 976 (2019).
15. Wang, Y., Lakoma, A. & Zogopoulos, G. Building towards Precision Oncology for Pancreatic Cancer: Real-World Challenges and Opportunities. doi:10.3390/genes11091098
16. Stathis, A. & Moore, M. J. Advanced pancreatic carcinoma: current treatment and future challenges. *Nat. Rev. Clin. Oncol.* **7**, 163–172 (2010).
17. Burris, H. A. *et al.* Improvements in survival and clinical benefit with gemcitabine as first- line therapy for patients with advanced pancreas cancer: A randomized trial. *J. Clin. Oncol.* **15**, 2403–2413 (1997).
18. Ciccolini, J., Serdjebi, C., Godefridus, ·, Peters, J. & Giovannetti, E. Pharmacokinetics and pharmacogenetics of Gemcitabine as a mainstay in adult and pediatric oncology: an EORTC-PAMM perspective. *Cancer Chemother. Pharmacol.* **78**, 1–12 (2016).
19. Deyme, L., Barbolosi, D. & Gattacceca, F. Population pharmacokinetics of FOLFIRINOX: a review of studies and parameters. *Cancer Chemotherapy and Pharmacology* **83**, 27–42 (2019).
20. Conroy, T. *et al.* FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer. *N. Engl. J. Med.* **364**, 1817–1825 (2011).

21. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nat. Med.* **17**, 500–503 (2011).
22. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. (2015). doi:10.1038/ng.3398
23. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).
24. Golan, T. *et al.* Maintenance Olaparib for Germline BRCA -Mutated Metastatic Pancreatic Cancer . *N. Engl. J. Med.* **381**, 317–327 (2019).
25. Marabelle, A. *et al.* Efficacy of pembrolizumab in patients with noncolorectal high microsatellite instability/ mismatch repair–deficient cancer: Results from the phase II KEYNOTE-158 study. *J. Clin. Oncol.* **38**, 1–10 (2020).
26. Hu, Z. I. *et al.* Evaluating mismatch repair deficiency in pancreatic adenocarcinoma: Challenges and recommendations. *Clin. Cancer Res.* **24**, 1326–1336 (2018).
27. Giardiello, F. M. *et al.* Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology* **119**, 1447–1453 (2000).
28. Wood, L. D., Yurgelun, M. B. & Goggins, M. G. Genetics of Familial and Sporadic Pancreatic Cancer. *Gastroenterology* **156**, 2041–2055 (2019).
29. Vasen, H. F. A. *et al.* Risk of developing pancreatic cancer in families with familial atypical multiple mole melanoma associated with a specific 19 deletion of p16 (p16-Leiden). *Int. J. Cancer* **87**, 809–811 (2000).
30. Lowenfels, A. B. *et al.* Hereditary pancreatitis and the risk of pancreatic cancer. *J. Natl. Cancer Inst.* **89**, 442–446 (1997).
31. Hu, C. *et al.* Association between inherited germline mutations in cancer predisposition genes and risk of pancreatic cancer. in *JAMA - Journal of the American Medical Association* **319**, 2401–2409 (American Medical Association, 2018).
32. Bartsch, D. K., Gress, T. M. & Langer, P. Familial pancreatic cancer-current knowledge. *Nature Reviews Gastroenterology and Hepatology* **9**, 445–453 (2012).
33. Klein, A. P. *et al.* Prospective risk of pancreatic cancer in familial pancreatic cancer kindreds. *Cancer Res.* **64**, 2634–8 (2004).
34. Humphris, J. L. *et al.* Clinical and pathologic features of familial pancreatic cancer. *Cancer* **120**, 3669–3675 (2014).
35. Klein, A. P. *et al.* Evidence for a major gene influencing risk of pancreatic cancer. *Genet. Epidemiol.* **23**, 133–149 (2002).
36. Canto, M. I. *et al.* Risk of Neoplastic Progression in Individuals at High Risk for Pancreatic Cancer Undergoing Long-term Surveillance. *Gastroenterology* **155**, 740-751.e2 (2018).
37. Vasen, H. *et al.* Benefit of surveillance for pancreatic cancer in high-risk individuals: Outcome of long-term prospective follow-up studies from three European expert centers. *J. Clin. Oncol.* **34**, 2010–2019 (2016).
38. Langer, P. *et al.* Five years of prospective screening of high-risk individuals from families with familial pancreatic cancer. *Gut* **58**, 1410–1418 (2009).
39. Canto, M. I. *et al.* Frequent detection of pancreatic lesions in asymptomatic high-

risk individuals. *Gastroenterology* **142**, 796–804 (2012).
40. Wong, C. *et al.* A region-based gene association study combined with a leave-one-out sensitivity analysis identifies SMG1 as a pancreatic cancer susceptibility gene. *PLoS Genet.*
41. Brumbaugh, K. M. *et al.* The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. *Mol. Cell* **14**, 585–598 (2004).
42. Brumbaugh, K. M. *et al.* The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. *Mol. Cell* **14**, 585–598 (2004).
43. Kirubakaran, S. Evolution of PIKK family kinase inhibitors A new age cancer therapeutics. *Front. Biosci.* **25**, 4866 (2020).
44. Imseng, S., Aylett, C. H. & Maier, T. Architecture and activation of phosphatidylinositol 3-kinase related kinases. *Current Opinion in Structural Biology* **49**, 177–189 (2018).
45. Lempiäinen, H. & Halazonetis, T. D. Emerging common themes in regulation of PIKKs and PI3Ks. *EMBO J.* **28**, 3067–73 (2009).
46. Morita, T. *et al.* Distant N- and C-terminal domains are required for intrinsic kinase activity of SMG-1, a critical component of nonsense-mediated mRNA decay. *J. Biol. Chem.* **282**, 7799–808 (2007).
47. Zhang, X. *et al.* SMG-1 inhibition by miR-192/-215 causes epithelial-mesenchymal transition in gastric carcinogenesis via activation of Wnt signaling. *Cancer Med.* **7**, 146–156 (2018).
48. Yin, Z. *et al.* LncRNA MAGI2-AS3 inhibits hepatocellular carcinoma cell proliferation and migration by targeting the miR-374b-5p/SMG1 signaling pathway. *J. Cell. Physiol.* **234**, 18825–18836 (2019).
49. Du, Y. *et al.* SMG1 acts as a novel potential tumor suppressor with epigenetic inactivation in acute myeloid leukemia. *Int. J. Mol. Sci.* **15**, 17065–17076 (2014).
50. Gubanova, E. *et al.* Downregulation of SMG-1 in HPV-positive head and neck squamous cell carcinoma due to promoter hypermethylation correlates with improved survival. *Clin. Cancer Res.* **18**, 1257–67 (2012).
51. Zhang, X. *et al.* SMG-1 inhibition by miR-192/-215 causes epithelial-mesenchymal transition in gastric carcinogenesis via activation of Wnt signaling. *Cancer Med.* **7**, 146–156 (2018).
52. Mai, S. J. *et al.* MicroRNA-18a promotes cancer progression through SMG1 suppression and mTOR pathway activation in nasopharyngeal carcinoma. *Cell Death Dis.* **10**, 819 (2019).
53. Han, L. L. *et al.* Expression and significance of the novel tumor-suppressor gene SMG-1 in hepatocellular carcinoma. *Oncol. Rep.* **31**, 2569–2578 (2014).
54. Du, Y. *et al.* SMG1 acts as a novel potential tumor suppressor with epigenetic inactivation in acute myeloid leukemia. *Int. J. Mol. Sci.* **15**, 17065–17076 (2014).
55. McIlwain, D. R. *et al.* Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12186–12191 (2010).
56. Ho, U. *et al.* SMG1 heterozygosity exacerbates haematopoietic cancer development in Atm null mice by increasing persistent DNA damage and oxidative

stress. *J. Cell. Mol. Med.* **23**, 8151–8160 (2019).
57. Gubanova, E., Issaeva, N., Gokturk, C., Djureinovic, T. & Helleday, T. SMG-1 suppresses CDK2 and tumor growth by regulating both the p53 and Cdc25A signaling pathways. *Cell Cycle* **12**, 3770–3780 (2013).
58. Denning, G., Jamieson, L., Maquat, L. E., Thompson, E. A. & Fields, A. P. Cloning of a novel phosphatidylinositol kinase-related kinase. Characterization of the human SMG-1 RNA surveillance protein. *J. Biol. Chem.* **276**, 22709–22714 (2001).
59. Yamashita, A., Ohnishi, T., Kashima, I., Taya, Y. & Ohno, S. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev.* **15**, 2215–2228 (2001).
60. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nature Reviews Molecular Cell Biology* **16**, 665–677 (2015).
61. Lykke-Andersen, S. *et al.* Human nonsense-mediated RNA decay initiates widely by endonucleolysis and targets snoRNA host genes. *Genes Dev.* **28**, 2498–2517 (2014).
62. Ni, J. Z. *et al.* Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**, 708–718 (2007).
63. Lareau, L. F., Inada, M., Green, R. E., Wengrod, J. C. & Brenner, S. E. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446**, 926–929 (2007).
64. Maquat, L. E., Tarn, W. Y. & Isken, O. The pioneer round of translation: Features and functions. *Cell* **142**, 368–374 (2010).
65. Brogna, S. & Wen, J. Nonsense-mediated mRNA decay (NMD) mechanisms. *Nature Structural and Molecular Biology* **16**, 107–113 (2009).
66. Popp, M. W.-L. & Maquat, L. E. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu. Rev. Genet.* **47**, 139–65 (2013).
67. Hug, N., Longman, D. & Cáceres, J. F. Mechanism and regulation of the nonsense-mediated decay pathway. *Nucleic Acids Res.* **44**, 1483–95 (2016).
68. Gubanova, E., Issaeva, N., Gokturk, C., Djureinovic, T. & Helleday, T. SMG-1 suppresses CDK2 and tumor growth by regulating both the p53 and Cdc25A signaling pathways. *Cell Cycle* **12**, 3770–3780 (2013).
69. Gewandter, J. S., Bambara, R. A. & O'Reilly, M. A. The RNA surveillance protein SMG1 activates p53 in response to DNA double-strand breaks but not exogenously oxidized mRNA. *Cell Cycle* **10**, 2561–7 (2011).
70. Gille, J. J. P. & Joenje, H. Chromosomal instability and progressive loss of chromosomes in HeLa cells during adaptation to hyperoxic growth conditions. *Mutat. Res. DNAging* **219**, 225–230 (1989).
71. Gehen, S. C., Staversky, R. J., Bambara, R. A., Keng, P. C. & O'Reilly, M. A. hSMG-1 and ATM sequentially and independently regulate the G1 checkpoint during oxidative stress. *Oncogene* **27**, 4065–4074 (2008).
72. Brown, J. A. L. *et al.* A Novel Role for hSMG-1 in Stress Granule Formation. *Mol. Cell. Biol.* **31**, 4417–4429 (2011).

73. Azzalin, C. M., Reichenbach, P., Khoriauli, L., Giulotto, E. & Lingner, J. Telomeric repeat-containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science (80-.).* **318**, 798–801 (2007).
74. Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: Progress, implications and challenges. *Hum. Mol. Genet.* **23**, (2014).
75. Kim, Y. G., Shi, Y., Berg, J. M. & Chandrasegaran, S. Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene* **203**, 43–49 (1997).
76. Gupta, D. *et al.* CRISPR-Cas9 system: A new-fangled dawn in gene editing. *Life Sciences* **232**, 116636 (2019).
77. Li, H. *et al.* Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects. *Signal Transduction and Targeted Therapy* **5**, 1–23 (2020).
78. Hryhorowicz, M., Lipiński, D., Zeyland, J. & Słomski, R. CRISPR/Cas9 Immune System as a Tool for Genome Engineering. *Arch. Immunol. Ther. Exp. (Warsz).* **65**, 233–240 (2017).
79. Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Molecular Cell* **58**, 568–574 (2015).
80. Maruyama, T. *et al.* Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.* **33**, 538–542 (2015).
81. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat. Protoc. 2013 811* **8**, 2180–2196 (2013).
82. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nat. 2014 5177536* **517**, 583–588 (2014).
83. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).
84. Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **34**, 339–344 (2016).
85. Aird, E. J., Lovendahl, K. N., St. Martin, A., Harris, R. S. & Gordon, W. R. Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* **1**, 54 (2018).
86. Kim, H. *et al.* A Co-CRISPR strategy for efficient genome editing in Caenorhabditis elegans. *Genetics* **197**, 1069–1080 (2014).
87. Agudelo, D. *et al.* Marker-free coselection for CRISPR-driven genome editing in human cells. *Nat. Methods* **14**, 615–620 (2017).
88. Shao, S. *et al.* Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing Saccharomyces cerevisiae Rad52. *Int. J. Biochem. Cell Biol.* **92**, 43–52 (2017).
89. Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766 (2014).
90. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).
91. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable

editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
92. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
93. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR–Cas9 screens. *Nat. 2019 5687753* **568**, 511–516 (2019).
94. Xu, S. *et al.* Genome-wide CRISPR screen identifies ELP5 as a determinant of gemcitabine sensitivity in gallbladder cancer. *Nat. Commun. 2019 101* **10**, 1–14 (2019).
95. Tang, Y. J. *et al.* Tracing Tumor Evolution in Sarcoma Reveals Clonal Origin of Advanced Metastasis. *Cell Rep.* **28**, 2837-2850.e5 (2019).
96. Rath, A. *et al.* Functional interrogation of Lynch syndrome-associated MSH2 missense variants via CRISPR-Cas9 gene editing in human embryonic stem cells. *Hum. Mutat.* **40**, 2044–2056 (2019).
97. Boettcher, S. *et al.* A dominant-negative effect drives selection of TP53 missense mutations in myeloid malignancies. *Science* **365**, 599 (2019).
98. S, C. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246–1260 (2015).
99. Moyer, C. L. *et al.* Rare BRIP1 missense alleles confer risk for ovarian and breast cancer. *Cancer Res.* **80**, 857–867 (2020).
100. Bolukbasi, M. F., Gupta, A. & Wolfe, S. A. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nature Methods* **13**, 41–50 (2015).
101. Zhang, F., Wen, Y. & Guo, X. CRISPR/Cas9 for genome editing: progress, implications and challenges. *Hum. Mol. Genet.* **23**, R40–R46 (2014).
102. McDonald, E. R. *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577-592.e10 (2017).
103. Gillet, J. P., Varma, S. & Gottesman, M. M. The Clinical Relevance of Cancer Cell Lines. *JNCI J. Natl. Cancer Inst.* **105**, 452–458 (2013).
104. Hoebeeck, J. *et al.* Rapid detection of VHL exon deletions using real-time quantitative PCR. *Lab. Investig.* **85**, 24–33 (2005).
105. Justus, C. R., Leffler, N., Ruiz-Echevarria, M. & Yang, L. V. In vitro Cell Migration and Invasion Assays. *J. Vis. Exp.* (2014). doi:10.3791/51046
106. Franken, N. A. P., Rodermond, H. M., Stap, J., Haveman, J. & van Bree, C. Clonogenic assay of cells in vitro. *Nat. Protoc.* **1**, 2315–2319 (2006).
107. Fukumoto, Y. *et al.* Cost-effective gene transfection by DNA compaction at pH 4.0 using acidified, long shelf-life polyethylenimine. *Cytotechnology* **62**, 73–82 (2010).
108. Brumbaugh, K. M. *et al.* The mRNA Surveillance Protein hSMG-1 Functions in Genotoxic Stress Response Pathways in Mammalian Cells. *Mol. Cell* **14**, 585–598 (2004).
109. Beigl, T. B., Kjosås, I., Seljeseth, E., Glomnes, N. & Aksnes, H. Efficient and crucial quality control of HAP1 cell ploidy status. *Biol. Open* **9**, (2020).
110. Kleeff, J., Korc, M., Apte, M., Vecchia, C. La & Johnson, C. D. Pancreatic cancer. *Nat. Publ. Gr.* **2**, 1–23 (2016).
111. Chhoda, A., Lu, L., Clerkin, B. M., Risch, H. & Farrell, J. J. Current Approaches to Pancreatic Cancer Screening. *American Journal of Pathology* **189**, 22–35 (2019).

112. Grant, R. C. *et al.* Exome-Wide Association Study of Pancreatic Cancer Risk. *Gastroenterology* **154**, 719-722.e3 (2018).
113. Smith, A. L. *et al.* Candidate DNA repair susceptibility genes identified by exome sequencing in high-risk pancreatic cancer. *Cancer Lett.* **370**, 302–312 (2016).
114. Hu, C. *et al.* Prevalence of pathogenic mutations in cancer predisposition genes among pancreatic cancer patients. *Cancer Epidemiol. Biomarkers Prev.* **25**, 207–211 (2016).
115. Roberts, T. L. *et al.* Smg1 haploinsufficiency predisposes to tumor formation and inflammation. *Proc. Natl. Acad. Sci.* **110**, E285–E294 (2013).
116. Hanlon, K. *et al.* Single-cell cloning of human T-cell lines reveals clonal variation in cell death responses to chemotherapeutics. *Cancer Genet.* **237**, 69 (2019).
117. Usuki, F. *et al.* Inhibition of SMG-8, a subunit of SMG-1 kinase, ameliorates nonsense-mediated mRNA decay-exacerbated mutant phenotypes without cytotoxicity. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15037–15042 (2013).
118. Evdokimova, V. *et al.* Translational Activation of Snail1 and Other Developmentally Regulated Transcription Factors by YB-1 Promotes an Epithelial-Mesenchymal Transition. *Cancer Cell* **15**, 402–415 (2009).
119. Evdokimova, V., Tognon, C., Ng, T. & Sorensen, P. H. B. Reduced proliferation and enhanced migration: Two sides of the same coin? Molecular mechanisms of metastatic progression by YB-1. *Cell Cycle* **8**, 2901–2906 (2009).
120. Yadav, D. & Lowenfels, A. B. The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology* **144**, 1252–1261 (2013).
121. Maresch, R. *et al.* Multiplexed pancreatic genome engineering and cancer induction by transfection-based CRISPR/Cas9 delivery in mice. *Nat. Commun.* **7**, (2016).
122. C, S. *et al.* The tumorigenicity diversification in human embryonic kidney 293 cell line cultured in vitro. *Biologicals* **36**, 263–268 (2008).
123. Fernandes, R., Nogueira, G., da Costa, P. J., Pinto, F. & RomÃ£o, L. Nonsense-Mediated mRNA Decay in Development, Stress and Cancer. in *Advances in Experimental Medicine and Biology* **1157**, 41–83 (Springer New York LLC, 2019).
124. Supek, F., Lehner, B. & Lindeboom, R. G. H. To NMD or Not To NMD: Nonsense-Mediated mRNA Decay in Cancer and Other Genetic Diseases. *Trends in Genetics* **37**, 657–668 (2021).
125. Hu, Z., Yau, C. & Ahmed, A. A. A pan-cancer genome-wide analysis reveals tumour dependencies by induction of nonsense-mediated decay. *Nat. Commun.* **8**, (2017).
126. Pastor, F., Kolonias, D., Giangrande, P. H. & Gilboa, E. Induction of tumour immunity by targeted inhibition of nonsense-mediated mRNA decay. *Nature* **465**, 227–230 (2010).
127. Lindeboom, R. G. H., Vermeulen, M., Lehner, B. & Supek, F. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat. Genet.* **51**, 1645–1651 (2019).
128. Litchfield, K. *et al.* Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat. Commun.* **11**, 1–11 (2020).
129. Boonen, R. A. C. M. *et al.* Functional analysis of genetic variants in the high-risk breast cancer susceptibility gene PALB2. *Nat. Commun.* **10**, 1–15 (2019).

130. Zhang, X. *et al.* Somatic superenhancer duplications and hotspot mutations lead to oncogenic activation of the KLF5 transcription factor. *Cancer Discov.* **8**, 108–125 (2018).