Investigating the Role of p53-Dependent Arresten in Tumour Progression and as a Cancer Biomarker

Serena Nath

Department of Biochemistry

McGill University

Montréal, Québec, Canada

August 2021

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

Table of Contents

List of Tables

List of Figures

List of Abbreviations

5-FU	5-fluorouracil
Ad-lacZ	adenoviral vector expressing lacZ
Ad-p53	adenoviral vector expressing p53
α(II)PH	alpha(II) prolyl 4-hydroxylase
bFGF	basic fibroblast growth factor
DMSO	dimethyl sulfoxide
CPT	camptothecin
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
ctRNP	crRNA + tracrRNA and Cas9 protein
DMEM	Dulbecco's Modified Eagle Medium
DSB	double stranded break
ECM	extracellular matrix
ELISA	enzyme-linked immunosorbent assay
ER	estrogen receptor
ES cells	embryonic stem cells
FAK	focal adhesion kinase
FBS	fetal bovine serum
gRNA	guide RNA
HDR	homology directed repair
IMPC	international mouse phenotyping consortium
indel	insertion/deletion mutation

HIF-1α	hypoxia-inducible factor 1-alpha
HRP	horseradish peroxidase
ITS	insulin-transferrin-selenium
ECL	enhanced chemiluminescence
IVF	*in vitro* fertilization
Mdm2	mouse double minute 2 homolog
MMP	matrix metalloproteinase
NC1	non-collagenous domain 1
NEHJ	non-homologous end joining
NSCLC	non-small cell lung carcinoma
PAM	protospacer adjacent motif
PCR	polymerase chain reaction
sgRNA	single gRNA
ssODN	single-stranded oligo donor
TAKE	Technique for Animal Knockout system by Electroporation
TAM	tamoxifen
TBS-T	tris(hydroxymethyl)aminomethane-buffered saline in 0.5% Tween 20
tracrRNA	trans-activating crRNA
TSP-1	thrombospondin-1
VEGF	vascular endothelial growth factor
WT	wild type

Abstract

Collagen-derived matrikines play a variety of roles, but specifically in regard to cancer, collagen-derived matrikines have been shown to have anti-angiogenic and anti-tumorigenic properties. Thus, these matrikines may play important roles in the development of cancer and can possibly be exploited for use in cancer diagnosis and cancer therapy. One collagen derived matrikine that has displayed promising anti-angiogenic and anti-tumorigenic characteristics is Arresten, a matrikine produced from the non-collagenous domain of the α1 chain of collagen IV. One-way the tumour suppressor p53 has been thought to modulate its tumour suppressor function is by activating the expression of Arresten. Therefore, in many cancers, where p53 function is compromised, Arresten expression may be inhibited, suggesting that Arresten may have applications as a biomarker for cancer detection. To investigate the role of Arresten in cancer detection, it was first confirmed that Arresten protein levels are a direct consequence of p53 action, and that secreted Arresten levels can be quantified by ELISA. Secreted Arresten levels in human plasma levels of non-small cell lung carcinoma patients were also quantified via ELISA. Processing of Arresten is highly unknown, but this information is essential when investigating the role of Arresten in cancer therapy. Therefore, cells were treated with the matrix-metalloproteinase inhibitors GM6001 and Marimastat to characterize the processing of Arresten. It was observed that Arresten is likely processed by a protease inhibited by GM6001 but not Marimastat. Additionally, to further investigate the role of Arresten in the tumour microenvironment for cancer therapy purposes, this study aimed to generate a mouse strain where the *COL4A1* gene is flanked by loxP sequences to eventually allow for conditional deletion of *COL4A1*, and consequently Arresten, by Cre recombinase. Although, this mutant mouse strain was not achieved in the end, it was confirmed via Sanger sequencing that the upstream loxP sequence is integrated in two founder mice, and the downstream loxP sequence is

possibly integrated into one of these two founder mice. Moreover, this study further optimized the process of using CRISPR-Cas9 to create conditional mouse models.

Résumé

Les matrikines dérivées du collagène avoir divers rôles, mais en ce qui concerne le cancer, il a été démontré que les matrikines dérivées du collagène ont des propriétés antiangiogenèseet anticancérigène. Ainsi, ces matrikines peuvent avoir rôle important dans le développement du cancer et peuvent éventuellement être utilisées pour le diagnostic et la thérapie du cancer. Une matrikine dérivée du collagène qui a montré des caractéristiques antiangiogenèse et anticancérigène prometteuses est Arresten, une matrikine produite à partir du domaine non-collagène de la chaîne α1 du collagène IV. On a pensé que le suppresseur de tumeur p53 modulait sa fonction de suppresseur de tumeur en activant l'expression d'Arresten. Par conséquent, dans de nombreux cancers, où la fonction p53 est compromise, l'expression Arresten peut être inhibée, suggérant qu'Arresten peut avoir des applications comme biomarqueur pour la détection du cancer. Pour étudier le rôle d'Arresten dans la détection du cancer, il a d'abord été confirmé que les niveaux de protéines Arresten sont une conséquence directe de l'action de p53, et que les niveaux d'Arresten sécrétés peuvent être quantifiés par ELISA. Les taux d'arrestations sécrétées dans le plasma humain des patients atteints de carcinome pulmonaire non à petites cellules ont également été quantifiés par ELISA. Le traitement de l'Arresten est très inconnu, mais cette information est essentielle pour étudier le rôle de l'Arresten dans le traitement du cancer. Par conséquent, les cellules ont été traitées avec les inhibiteurs de la métalloprotéinase matricielle GM6001 et Marimastat pour caractériser le traitement d'Arresten. On a observé qu'Arresten est probablement traité par une protéase inhibée par GM6001, mais pas par Marimastat. De plus, pour approfondir l'étude du rôle d'Arresten dans le microenvironnement tumoral à des fins de traitement du cancer, cette étude visait à générer une souche de souris où le gène *COL4A1* est flanqué de séquences loxP pour permettre à terme la délétion conditionnelle de *COL4A1*, et par conséquent Arresten, par Cre recombinase.

Bien que cette souche mutante de souris n'ait pas été atteinte à la fin, il a été confirmé par séquençage Sanger que la séquence loxP en amont est intégrée dans deux souris fondatrices et la séquence loxP en aval est éventuellement intégrée dans l'une de ces deux souris fondatrices. De plus, cette étude a optimisé le processus d'utilisation de CRISPR-Cas9 pour créer des modèles de souris conditionnels.

Acknowledgments

I would like to thank Dr. Jose Teodoro for providing me with the opportunity to work in his lab and for the chance to work on this exciting project. His mentorship, guidance, and support were invaluable, and I learned a lot from him. His kindness and patience were also unparalleled, and I cannot thank him enough for taking me as a graduate student. I would also like to thank the best research assistant in the universe, Isabelle Gamache. I cannot thank her enough for her guidance with troubleshooting my experiments, with discussing new ideas for my project, and with helping me develop into a better scientist, but also for life lessons in perseverance and determination. I would also especially like to thank her for her support and confidence in my project, especially when I did not have confidence. Thank you to my research advisory council members, Dr. Peter Siegel, and Dr. Jerry Pelletier, who provided valuable insight and guidance for my project. Thank you also to the McGill Transgenic core for their help in trying to create our floxed mouse strain, and for their help with the *in-vitro* Cre assay.

To my labmates, thank you all for all of your friendship and support throughout these past few years. I am so grateful for all of your help with troubleshooting experiments, even though we all worked on different projects, for the support when I encountered difficulties in my project, and all of the laughter and fun conversation. I would like to give a special thanks to Alisha, Yilin, Owen, Mohamed, and Cynthia for all of your friendship and fun conversations in our "picnic" group chat. I would also like to thank my friends Marina, Rachel, Alyssa, and Chinchin for their support, love, and fun adventures in Montreal. To Lauren, I am so thankful for your friendship, love, and support during my thesis, and for your visits to Montreal. To my mother Sanu, father Parveen, and sister Aneesha, thank you for all of your support and love. Special thanks to my mom for always listening to me discuss my project and the day-to-day ongoings of

the lab, even if she did not understand my project. However, I hope she now knows what CRISPR is. Special thanks also to my dad for always encouraging me to purse graduate school, and to my sister for always making me laugh. Finally, I would like to thank McGill University, which is situated on the territory of the Kanien'Kehà:ka, for providing me with the opportunity to study at the Goodman Cancer Research Centre.

Preface

The author of this thesis performed all experiments to screen Arresten as a biomarker for cancer detection using ELISA and all experiments to investigate the processing of Arresten. The author of this thesis also designed all primers, CRISPR sgRNAs, and repair templates, performed all experiments to screen blastocysts, founder mice, and F1 generation mice, analyzed all data from Sanger sequencing, and created all tables and figures contained in this thesis, except for figure 7D which was created by the McGill Transgenic Core. Additionally, the McGill Transgenic Core contributed the following to this thesis: IVF to generate C57BL/6J zygotes, microinjection, and electroporation of CRISPR-Cas9 components into C57BL/6J zygotes, extraction of DNA from blastocysts, and initial housing and care of the animals.

1. INTRODUCTION

1.1 The Tumour Microenvironment

The tumour microenvironment is the environment surrounding the tumour consisting of extracellular matrix (ECM), as well as myofibroblasts, fibroblasts, neuroendocrine cells, adipose cells, immune-inflammatory cells, the blood and lymphatic vascular networks and more, all of which interact with the tumour to influence tumour progression. The tumour microenvironment influences cancer progression by influencing major processes such as evasion from growth suppressors, promoting invasion and metastasis, resisting apoptosis, stimulating angiogenesis, maintaining proliferative signaling, evading immune destruction, genome instability and mutation [1].

1.2 Matrikines

One of the most abundant components of the tumour microenvironment is the ECM. The ECM is a dynamic entity that undergoes frequent remodeling based on cellular signals that it receives [2]. This frequent remodeling leads to the liberation of matrikines**,** which are peptides that have been liberated from ECM macromolecules and are able to regulate cell activities such as cell adhesion, migration, proliferation, protein synthesis, and apoptosis [3]. Many of these parent macromolecules have been shown to produce many different matrikines, with the functional activity of each matrikine usually differing from the function of the parent protein. Matrikines are either classified as “natural” matrikines or “cryptic” matrikines [4]. Natural matrikines typically have naturally occurring domains within the parent protein that do not require proteolysis to exert their function, whereas cryptic matrikines typically have hidden domains within the parent protein that require liberation via proteolysis from the parent protein to become active. In the context of cancer, these matrikines are able to influence the proliferation

and invasive properties of tumor or inflammatory cells, and the angiogenic properties of the tumour [5].

1.3 Overview of Collagen

Collagen is the most abundant protein in mammals [6]. A major ECM macromolecule is collagen, which plays a structural support role for cells and tissues, and an instructive role for the cellular microenvironment. Currently, there are 28 different types of collagens in vertebrates which are classified as either fibril or non-fibril forming [7]. However, most collagen is collagen type 1, a fibril forming collagen. The general role of collagen depends on the type of collagen. Collagens 1-3, 5 11, 24, and 27 are fibril collagens and thus make up part of the interstitial matrix, whereas collagens 4, 15, and 18 are non-fibril and compose part of basement membranes [8], [9], [10].

Mature collagens are composed of three left-handed alpha chains, which are polypeptide chains that form a right-handed super helix [7] . Each alpha chain contains at least one triple-helical domain, which is present in all collagens. But depending on the type of collagen, the percentage of the collagen that is composed of the triple helix varies [7]. The typical pattern of the triple helix motif, also known as the collagenous domain, of each alpha chain is Gly-X-Y, where X is usually proline and Y is usually 4-hydroxyproline [11]. The glycine in each alpha chain allows for the trimeric alpha chains to closely associate with each other, while the proline and hydroxyproline allow for the alpha chain to associate with other components of the ECM and for the tight, left-hand twisting structure of the alpha chains [12]. Flanking the collagenous domain of the alpha chains is usually the N-propeptide and C-propeptide domains on the N-terminus and C-terminus sides, respectively [11]. Folding of the triple helix domain is initiated by the trimerization domain, which differs depending on the type of collagen [13]. For fibrillar

collagens, formation is initiated by the C-terminus, but for non-fibrillar collagens, formation is initiated by the N-terminus. Collagens differ from each other due to differences of their distinct domains, differences in their individual alpha chains and the different combinations of their trimeric alpha chain associations, which can either be heteromeric or homomeric. Fibrillar collagens are composed of an uninterrupted collagenous domain, whereas non-fibrillar collagens have interruptions in their collagenous domains [7].

Collagens play many roles in vertebrates. In particular, fibril collagens provide structural support in vertebrates, which is achieved by helping support molecular architecture, shape and mechanical properties of tissues [7]. Other collagens play a role in supporting tissue integrity. For example, excess collagens which are secreted into the ECM to limit fibrosis have a less defined structure, compared to fibril collagens, and thus can interact with cellular receptors to modulate cellular processes [14]. Moreover, certain collagens have more specialized roles too, such as membrane collagen, collagen IV which has been shown to play a role in the development of the vertebrate nervous system, and other collagens which have been found in the brains of patients with Alzheimer's diseases and are therefore thought to play a role in Alzheimer's [15], [16].

Collagens also play a role through the action of matrikines derived from collagen. Turnover of collagen leads to the production of matrikines derived from the collagenous and non-collagenous domains [17]. These matrikines have been shown to regulate many different processes such as angiogenesis, tissue repair, tumour growth, and tissue development. Multiple matrikines can be derived from the same parent collagen, but matrikines are usually derived from basement membrane collagens [18]. Many of these collagens have shown anti-angiogenic and anti-tumourigenic properties, and therefore have been investigated as potential cancer therapies.

One of the most famous examples of a collagen derived matrikine being investigated as a cancer therapy is endostatin, a matrikine derived from the C-terminal of the non-collagenous (NC1) domain of collagen XVIII [19]. In 1997, Endostatin was discovered in a murine hemangioendothelioma cell line, and subsequent research demonstrated its robust anti-angiogenic effect via upregulation of various anti-angiogenic factors and down-regulation of various pro-angiogenic factors. As a result, clinical studies were performed to investigate the ability of Endostatin to act as a cancer therapy by inhibiting tumour angiogenesis and growth [20]. Unfortunately, initial clinical trials with Endostatin were unsuccessful, due to the ineffectiveness of Endostatin and its short-half life. However, Endostar, a modified recombinant human Endostatin has shown success in clinical trials and has been approved for treatment of non-small cell lung carcinoma (NSCLC) in China [21]. Another, not as well characterized and investigated, example is Arresten, a matrikine derived from the α1 chain of collagen IV, which has been shown to have anti-angiogenic and anti-tumorigenic activity [22].

There are two more other well-known collagen derived matrikines being investigated as biomarkers; endostatin and tumstatin. Tumstatin is derived from the α3 chain of collagen IV. Studies have investigated endostatin as a prognostic biomarker for acute ischemic stroke and as a marker for cardiovascular health, such as vascular and myocardial damage and prognosis for different cardiovascular events [23], [24]. Tumstatin has been investigated as a diagnostic biomarker for various lung disorders [25]. Collagen IV has also been investigated as a biomarker for pancreatic cancer, where circulating levels of type IV collagen were measured by enzyme-linked immunosorbet assay (ELISA) in patients with pancreatic cancer, and it was shown that patients with pancreatic cancer have significantly higher circulating levels of type IV collagen compared to control patients [26].

1.4 Type IV Collagen

Type IV collagen is a major component of basement membranes, and acts in a structural role to support basement membranes by providing a scaffold that binds laminin, fibronectin, enactin, and proteoglycans [8], [9]. It is formed from different combinations of three out of six different α chains (α1- α6) that are encoded by six different genes (*COL4A1-COL4A6*). All of the six genes encoding the six α chains are organized in pairs in a head-to head arrangement, in which the gene pair shares a common promoter [3]. The *COL4A1* gene is arranged in a head-to-head arrangement with the *COL4A2* on chromosome 13 in humans; the *COL4A3* gene is arranged in a head-to-head arrangement with the *COL4A4* gene on chromosome 2 in humans; and the *COL4A5* gene is arranged in a head-to-head arrangement with the *COL4A6* gene on chromosome 22 in humans [22],[27],[28]. The different alpha chain trimers that exist for type IV collagen are $\alpha1_2\alpha2$, α3α4α5, and $\alpha5_2\alpha6$, with the distribution of each trimer differing [22]. The most ubiquitous trimer is the $\alpha1_2\alpha2$ which is widespread through various tissues. However, the α3α4α5 trimer is only found in basement membranes of the aorta, the pulmonary alveoli, the cochlea, the glomerulus, and the lens capsule, whereas the $\alpha5_2\alpha6$ trimer is only found in basement membranes in the kidney, lung, esophagus, and skin. Each alpha chain is composed of three domains: an N-terminal 7s domain, middle triple-helical domain with several interruptions, and C-terminal globular NC1 domain [3]. Collagen IV has been shown to produce several matrikines from its different α chains, all which have shown to influence the angiogenesis process, albeit via different mechanisms. From the NC1 domain of the α3 chain, it has been shown that the anti-angiogenic matrikine Tumstatin is produced, from the NC1 domain of the α2 chain, the anti-angiogenic matrikine canstatin is produced and lastly, the NC1 domain of the α1 chain is processed to produce the anti-angiogenic and anti-tumorigenic matrikine Arresten [29].

1.5 Angiogenesis

Angiogenesis is the process by which new blood vessels are formed from pre-existing blood vessels [30]. It is a natural process that occurs during various events such as wound healing, embryonic development, the menstrual cycle in biological females, and endometrium remodeling [31]. However, this process is also essential for tumour tumours to continue to grow, as it is allows for continued delivery of nutrients and oxygen required for the cells of the tumour to survive. Without their own blood supply, solid tumours cannot grow beyond 2 to 3 mm in diameter, and instead may undergo apoptosis [32]. Additionally, angiogenesis is required at two stages for the tumour to metastasize. Firstly, angiogenesis is required for metastatic tumour cells to be shed from the primary tumour. Secondly, angiogenesis is required once the metastatic tumour cells arrive at their secondary target organ [31].

In order for angiogenesis to occur in tumours the angiogenic switch must be induced, meaning that the angiogenic phenotype has been induced in these cells. This switch is driven by a multitude of factors [31]. Firstly, angiogenic oncogenes can influence the balance of angiogenic factors by upregulating the expression of pro-angiogenic factors and down regulating the expression of anti-angiogenic factors. Pro-angiogenic factors include basic fibroblast growth factor (bFGF), vascular endothelial growth factor (VEGF), and angiogenin, which are released from tumour cells, the ECM, and by macrophages attached to the tumour, whereas anti-angiogenic factors include thrombospondin-1 (TSP-1) [33]. Secondly, hypoxic conditions in the tumour can activate hypoxia-inducible factor 1-alpha (HIF-1α) which goes on to activate pro-angiogenic factors [34]. Thirdly, the tumor can induce fibroblasts to secrete pro-angiogenic factors [35]. Lastly, progenitor endothelial cells derived from bone marrow can induce angiogenesis when they are recruited to tumours [36].

1.6 Arresten

The anti-angiogenic and anti-tumorigenic matrikine Arresten is a 26 kDa peptide that is processed from the NC1 domain of the α1 chain of collagen IV [37]. Several studies have attempted to characterize the anti-angiogenic and anti-tumorigenic effects of *in vitro* and *in vivo.* While characterizing the role of Arresten, Colorado et al. (2000), demonstrated that Arresten is able to inhibit endothelial cell proliferation, migration, and tube formation *in vitro* [37]. They also demonstrated via a Matrigel plug assay in mice that Arresten can inhibit formation of new capillaries *in vivo* and that Arresten likely inhibits formation of new blood vessels by inhibiting multiple steps in the angiogenic process. Moreover, they were able to demonstrate that Arresten can inhibit the growth of xenografted human tumours in nude mice and the development of metastasis.

Then, to further understand the anti-angiogenic properties of Arresten, Nyberg et al. (2008), demonstrated using an annexin V-fluorescein isothiocyanate assay that Arresten is able to induce apoptosis of only epithelial cells *in vitro*. They then demonstrated using a TUNEL assay that Arresten is able to induce apoptosis in epithelial and tumour cells *in vivo* in mice, however it was proposed that the apoptosis of the tumour cells was due to limiting of the blood supply to the tumour cells due to apoptosis of the endothelial cells. This study also demonstrated using Arresten deletion mutants that the active site of Arresten is within the last 113 amino acids of Arresten, and α1β1 integrin is a functionally relevant receptor for Arresten to mediate its anti-angiogenic and anti-tumorigenic activity. Arresten has also been shown to influence other cells in the tumour microenvironment. It has been demonstrated that Arresten can also inhibit tumour growth by inhibiting proliferation, migration, and invasion of carcinoma cells via induction of

apoptosis of carcinoma cells [38]. Furthermore, Arresten is able to promote mesenchymal to endothelial transition of carcinoma cells.

It is postulated that arresten mediates its anti-angiogenic and anti-tumourigenic properties by getting cleaved at the cell surface and being released into the vasculature. Arresten then binds to α1β1 integrin on endothelial cells which leads to the inhibition of phosphorylation of focal adhesion kinase (FAK) [37]. Inhibiting the phosphorylation of FAK prevents leads to the inhibition of the Raf/MEK/ERK1/2/p38 mitogen-activated protein kinase signaling pathway, which ultimately causes decreases in the expression of HIF-1α and vascular endothelial growth factor, which are both pro-angiogenic factors. Decreased expression of these pro-angiogenic factors then results in inhibited endothelial cell migration, proliferation, and tube formation. Arresten also mediates its anti-angiogenic and anti-tumourigenic properties increasing apoptosis of endothelial cells by down-regulating anti-apoptotic molecules such as B-cell lymphoma 2 and B-cell lymphoma-extra large [39]. Arresten is also able to induce apoptosis via inhibition of p38-mitogen activated protein kinase signaling which results in activation of caspase-3/poly (ADP-ribose) polymerase cleavage [40].

1.7 The Tumour Suppressor p53

1.7.1 Overview of p53

The p53 protein is a highly important, multi-functional transcription factor, that regulates many different cellular processes. These processes include cell-cycle, DNA replication, uncontrolled cell division during tumour growth, senescence, apoptosis, and cell differentiation [41]**.** Studies have also suggested that p53 may play a role in allowing the cell to adjust its metabolism in response to mild physiological fluctuations [42]. However, p53 is also termed a

tumour suppressor, as it also regulates many cellular processes, such as protecting cells from uncontrolled cell growth, to prevent cancer formation [41]. Additionally, p53 has been implicated in inhibiting tumour growth via inhibition of metastasis and angiogenesis via three main mechanisms [43]. Firstly, through several direct and indirect mechanisms, p53 inhibits the production of various pro-angiogenic factors such as VEGF, bFGF**,** bFGF-binding protein, and cyclooxygenase-2 [44]. Furthermore, it has been shown that p53 can inhibit tumour growth via upregulation of anti-angiogenic factors by binding directly to the p53-response element, a DNA sequence motif, in the promoter of target genes. Examples of anti-angiogenic factors directly upregulated by p53 include TSP-1, brain specific-angiogenesis inhibitor 1, ephrin receptor 2, an anti-angiogenic collagen derived matrikines [43]. Lastly, under extreme hypoxic conditions, p53 inhibits a main regulator of the hypoxia response. Studies have shown that p53 directly binds HIF-1α and targets it for degradation [45]. As a result, HIF-1α, in conjunction with HIF-1β, is unable to bind hypoxic responsive elements in the promoters and enhancers of target genes, thus preventing the eventual activation of the angiogenesis pathway [46].

Under normal conditions of low cellular stress, p53 expression is maintained at a low level via the p53-mouse double minute 2 homolog (Mdm2) negative feedback loop which leads to the ubiquitination and degradation of p53 by the E3 ubiquitin Mdm2 [47]. When the cell experiences stress, such as with uncontrolled cell division and proliferation, DNA damage, hypoxia, or mitotic apparatus dysfunction, p53 is post-translationally modified resulting in the disruption of the p53-Mdm2 interaction. As a result, p53 levels stabilize and increase in the cell. This allows for p53 to activate and repress various target genes to deal with the cellular stress. Severe stress will eventually lead to p53 triggering apoptosis or senescence of the cell, whereas less extreme stress will lead to p53 briefly arresting growth of the cell to attempt to repair any damage and

deal with the stress [48]. But, in approximately half of all cancers p53 is mutated, and it is thought that in the other half of all cancers p53 function is inactivated due to the overexpression of Mdm2 [49]. Therefore, in most cancers, the function of p53 is compromised, including its ability to inhibit tumour growth.

1.7.2 Arresten and p53

One-way p53 modulates its tumour growth inhibition function is by activating the expression of Arresten. Studies have demonstrated that p53 can increase arresten production via three mechanisms [50]. Firstly, a genome wide analysis was performed for p53 binding sites. One of the strongest hits of this screen was a p53-binding site that was approximately 26 kbp downstream of the 3' end of the *COL4A1* gene. Using chromatin immunoprecipitation, it was then proposed that p53 interacts with the *COL4A1* promoter via long-rage looping mechanisms to increase transcription of the *COL4A1* gene [50]. Secondly, it was shown that p53 transcriptionally activates the expression of the enzyme α(II) prolyl 4-hydroxylase [α(II)PH], an enzyme required for collagen production. To investigate whether α(II)PH had any effect on collagen production, α(II)PH was overexpressed in cells. This resulted in an increase in full-length α1 collagen IV in cells and an increase in processed Arresten in the conditioned media. Thus, increasing the expression of this α(II)PH increases the amount of α1 collagen IV produced, and by extension, the amount of Arresten produced [50], [51]. Lastly, by examining the effect of p53 on the type IV collagen matrix produced by tumour cells, it was shown that p53 transcriptionally activates ECM proteases [50]. These proteases remodel the type IV collagen matrix by degrading this matrix, resulting in a destabilization of the basement membrane, and thus inhibition of the angiogenic process. Additionally, degradation of the collagen IV matrix would also result in an increased liberation of Arresten. Although it is unknown exactly which

ECM proteases process Arresten by remodeling the type IV collagen matrix, it has been hypothesized that this remodeling is due to matrix-metalloproteinases (MMPs). A previous study has demonstrated that p53 transcriptionally activates the MMP-2, an enzyme known to degrade type IV collagen [52]. Furthermore, studies have shown that the processing of other collagen derived matrikines is due to processing by matrix metalloproteinases. For example, Tumstatin, is processed from the α3 chain of collagen IV via proteolysis by MMP-9 [53].

1.8. *COL4A1* Knockout

To investigate the role of the collagen IV $\alpha 1_2 \alpha 2$ isoform in basement membrane assembly and embryonic development, Pöschl et al. (2004), created a *COL4A1* and *Col4A2* knockout in mice [54]. This knockout was achieved by deleting the first exon of *COL4A1*, the promoter of *COL4A1/2*, and exons 1-3 of *COL4A2* via replacement with a neomycin cassette. They found that lethality of the embryos of these mice occurred around day 10.5-11.5 of the embryonic stage due to impairment of basement membrane stability. Thus, this study shows that to study perform *in vivo* knockout studies of *COL4A1* and therefore Arresten, an alternative method, such as a conditional knockout, must be used.

1.9 Conditional Knockout Mouse Models

1.9.1 Overview of Conditional Knockout Mouse Models

The analysis of gene mutation *in vivo* presents an essential opportunity to gain further understanding of the gene and its applications in disease. When targeted mutagenesis of mouse germlines was first attempted, these mutations modified a gene in all cells, at all developmental stages of the mouse [55]. However, classic gene knockouts sometimes presented challenges for studying genes *in vivo.* Some germline mutations are lethal to the organism, and thus any attempt

to knockout certain genes would not produce a viable organism to study [56]. Furthermore, some genes are expressed only in specific cell types, or at a certain developmental stage [56]. Therefore, classic gene knockout models could prevent identification of the exact type of cell or the exact stage at which the protein plays a critical role. Lastly, knockout models make it difficult to study somatically acquired genetic diseases, and instead, are more optimal to study inherited genetic diseases [56].

To circumvent the issues observed with classical gene knockout models, the strategy of conditionally targeting genes in models was developed. This strategy takes advantage of the Cre-loxP bacteriophage P1 system to inactivate a target gene at a specific developmental point, and/or in a specific cell type [57], [58]. In this system, the Cre recombinase enzyme recognizes a 34 base pair sequence motif, known as a loxP sequence [59]. If a DNA sequence is between two loxP sequences in the same orientation, Cre will mediate recombination between these loxP sites to excise the DNA in between. If a DNA sequence is between two loxP sequences of opposite orientation, Cre will invert the DNA [60]. The region of DNA in between the loxP sequences is known as the critical exon. The critical exon is selected by having the following two criteria: 1) deletion will result in a frameshift of the reading frame leading to a truncated or non-functional protein; and 2) it must be present in all mRNA transcript isoforms [61].

To generate a conditional knockout mouse strain, two different strains of mice must be produced and mated [55]. The first strain is termed the floxed strain, where homologous recombination is exploited to introduce two loxP sequences to flank the critical exon. The second strain is the Cre expressing strain, where the Cre recombinase gene is expressed under a tissue specific or whole-body promoter. The two strains are then crossed to produce the conditional knockout strain. In the conditional knockout strain the Cre enzyme must be induced in order for

the knockout to occur. Typically, this is achieved by having the CRE recombinase fused to a mutant form of the human estrogen receptor (ER) ligand binding domain [62]. The most used mutant Cre-ER complex is the Cre-ER^{T2} complex, in which the human ER ligand binding domain contains three mutations: G400V/M543A/L544A [63]. Under physiological conditions, Cre-ER is sequestered in the cytoplasm by heat shock protein 90 (HSP90). Additionally, at physiological concentrations, mutations to the ER ligand binding domain prevents this protein complex from being sensitive to the natural ligand of the ER, 17β-estradiol [64]. However, it is highly sensitive to synthetic estrogen antagonists like tamoxifen (TAM) [65]. Therefore, to induce Cre activity in the mouse, the mouse is injected with TAM. TAM will bind the Cre-ER complex, allowing for the complex to dissociated from HSP90 and translocate into the nucleus, where it can mediate recombination between the loxP sites, thus producing the spatially and temporally controlled knockout.

1.9.2 Traditional Approaches for Creating Conditional Knockout Mouse Models

Historically, to create a conditional knockout mouse model, approaches to obtain conditional alleles in mouse embryonic stem (ES) cells were used [55]. With this approach, molecular cloning methods are used to assemble a targeting vector that contains a 5'homology region, the loxP-flanked region, a loxP-flanked antibiotic resistance gene, typically neomycin, to act as a positive selection marker on the opposite side of the floxed critical exon, and a 3' homology region [61]. In this targeting vector design, the floxed critical exon and floxed antibiotic resistance gene share one loxP sequence, so that in total there are three loxP sequences [55]. The targeting vector is then electroporated into ES cells, where the floxed critical exon and floxed antibiotic resistance gene are introduced into the target site via homologous recombination. ES cells carrying the targeting vector are then selected for via antibiotic selection [55]. Once the

positive ES clones have been identified, Cre recombinase is transiently introduced into the ES cells via transfection to remove the antibiotic resistance gene [55]. When Cre mediates recombination, there are three possible scenarios that can arise: 1) only the antibiotic resistance gene is excised while the floxed critical exon remains; 2) the floxed critical exon is excised while the antibiotic resistance gene remains; and 3) both the antibiotic resistance gene and floxed critical exon are excised [55]. To select for the ES clones that have only the antibiotic resistance gene and not the critical exon excised, southern blot or polymerase chain reaction (PCR) is typically used [55]. Once the desired ES clones have been identified, they are microinjected into blastocysts. These blastocysts are then transferred to pseudo pregnant female mice, from which germline chimeric mice are born [55]. Mice containing the floxed critical exon are then mated with a Cre expressing mouse strain to produce the conditional knockout mouse strain. As it became more popular to create conditional knockout mouse models, the International Mouse Phenotyping Consortium (IMPC) was launched to aid in the process of creating conditional knockout mouse models [61]. This program is a public library of conditionally targeted gene vectors in mouse ES cells, conditional mutant ES cells, and conditional mutant mice, all which can be bought and used by researchers.

1.9.3 Use of CRISPR for Generation of Conditional Knockout Mouse Models

The process of using ES cells to generate conditional knockout mouse models is time-consuming, laborious, expensive, and inefficient, even with the additional help from the IMPC. Therefore, alternative methods for generating conditional knockout mouse models have been explored, most notably the Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) system which has revolutionized the process of generating conditional knockout mouse models. CRISPR DNA sequences were first observed by Ishino et al. (1987) in *Escherichia coli*;

however, at this point, the sequences did not have the name of CRISPR [66]. Further analysis of CRISPR DNA sequences was performed during the human genome project, where researchers identified key features of CRISPR repeat and spacer elements [67]. This then led many researchers to investigate CRISPR sequences. Eventually, further research revealed more insight regarding the function of CRISPR systems. In particular, Barrangou et al. (2007) demonstrated that CRISPR provides a defense against viruses in prokaryotes [68]. Furthermore, research by several labs elucidated the mechanism of action of CRISPR systems and further characterized the individual components of the CRISPR system [67].

To date, there are three different types of CRISPR-Cas systems that have been identified, in which the three systems mainly differ in regards to how they produce CRISPR RNA (crRNA) and Cas proteins [69]**.** However, generally, the use of all CRISPR systems as a defence mechanism in prokaryotes occurs in two stages: the immunization stage, and the immunity stage [70]. During the immunization stage, when a novel virus infects a prokaryote, fragments of viral DNA, termed spacer elements, are integrated within the prokaryotic host CRISPR loci. The CRISPR loci are cluster of spacer elements that are interspersed by the CRISPR DNA sequences. During the immunity stage, when the same virus infects, the spacer element is transcribed to produce pre-processed crRNA and then modified by Cas proteins, which are also expressed during this stage, and host factors, to produce mature crRNA. The crRNA then forms a complex with a Cas RNA-guided nuclease, in which the crRNA acts as an antisense guide for the Cas protein. This complex is then able to bind foreign viral DNA sequences and cleave this DNA.

Scientists have been able to adapt the CRISPR system for extensive use in genome editing, thus largely replacing the previously used zinc-finger nucleases and transcription activator-like effector nucleases [67]. The type II CRISPR-Cas9 system from *Streptococcus pyogenes* is the

most widely used system for genome editing due to its simple NGG protospacer adjacent motif (PAM) sequence requirements [67]**.** Specific to the CRISPR-Cas9 system is the requirement of a short RNA, termed the trans-activating crRNA (tracrRNA) in addition to the crRNA, and for the target DNA to immediately precede the NGG PAM [71],[72],[73]. In this system, the crRNA binds to the tracrRNA, forming the functional guide RNA (gRNA). The crRNA portion of the gRNA still acts as a guide to direct the gRNA-Cas9 effector complex to the matching viral DNA sequence, whereas the tracrRNA aids the association of the gRNA with Cas9. To adapt the CRISPR system for use in genome editing, Jinek et al. (2012) demonstrated that the gRNA composed of crRNA and tracrRNA could be engineered as a single gRNA (sgRNA) chimera [74]. Furthermore, other studies demonstrated that the CRISPR-Cas9 system could adapted to be used for genome editing in eukaryotic cells. These adapted CRISPR-Cas9 systems allow for sgRNA to be designed to target any DNA sequence in the desired genome, provided that DNA sequence immediately precedes NGG PAM, to induce a double stranded break (DSB) in the DNA at this area. A DSB is then repaired either through one of two mechanisms. First, the DSB could be repaired via the non-homologous end-joining pathway (NHEJ) which would leave insertion/deletion (indel) mutations in the DNA [75]. Therefore, this repair pathway provides a way to mediate gene knockouts if the indel occurs within an exon of the gene and results in a frameshift mutation or premature stop codon. The second mechanism by which a DSB can be repaired is via homology directed repair (HDR). If an exogenously produced repair template is provided to the cell, the HDR pathway can be leveraged to introduce new DNA sequences into the location where the DSB occurred [76].

The type II CRISPR-Cas9 system has been adapted for use in creating genetically modified mouse models, including conditional knockout mouse models. A method, termed EASI-CRISPR,

developed by Miura et al. (2017) demonstrated the use of CRISPR-Cas9 for creating knock-in and conditional knockout mouse models [77]. To create the floxed strain which will be mated with the Cre expressing strain to produce the conditional knockout model, mouse zygotes are injected with two pre-assembled crRNA + tracrRNA and Cas9 protein (ctRNP) complex. Each ctRNP complex is used to target one of the regions flanking the critical exon where a loxP sequence should be inserted. The repair template, contains the floxed critical exon which is flanked by homology arms of about 200 base pairs, allowing for the repair template to be inserted at the Cas9 cleavage sites via HDR. The mouse zygotes are then implanted into a pseudo pregnant female mouse who will give birth to founder pups for the floxed strain.

1.10 Research Objectives

The research objectives of this thesis were as follows:

1. Investigate Arresten for use as a biomarker for cancer detection
2. Identify proteases responsible for the processing of Arresten from COL4A1
3. Use the CRISPR-Cas9 system to create a floxed *COL4A1* mouse strain

Previous unpublished work in our lab has suggested that Arresten levels may be lower in the plasma of patients with NSCLC and pancreatic cancer compared to healthy individuals. Thus, these results suggest that Arresten could have potential use as a biomarker for cancer detection. Characterizing the effect of p53 on Arresten protein levels, and further investigating Arresten levels in NSCLC patients, could help further establish the relevance of Arresten as a biomarker for cancer detection. As mentioned earlier, the proteases responsible for the processing of Arresten from the COL4A1 is highly unknown. However, further investigation and characterization of the processing of Arresten from the ECM would provide valuable information

for eventual investigation of Arresten as a cancer therapeutic. Furthermore, if MMPs are truly responsible for processing of Arresten, then determining which proteases are responsible could also provide valuable information for the use of MMP inhibitors in cancer therapy. MMP inhibitors have been investigated in cancer therapy because of evidence suggesting that MMPs play a role in promoting tumour progression and metastasis [78]. However, these inhibitors were unsuccessful in clinical trials, possibly because MMPs may also have anti-angiogenic and anti-tumourigenic roles by releasing anti-angiogenic and anti-tumourigenic matrikines from the ECM. Thus, characterizing the processing of matrikines, including Arresten, by MMPs, could aid in developing new, more specific MMP inhibitors that do not affect matrikine processing, but rather only affect the role of MMPs in promoting tumour progression. Additionally, creating a floxed *COL4A1* mouse strain would allow for eventual conditional knockout of *COL4A1*, and by extension Arresten, when this strain is mated with a Cre expressing mouse strain. As a result, this would allow for investigation of the *in vivo* effects of Arresten on the tumour microenvironment, which has previously never been investigated. Further characterization of Arresten on the tumour microenvironment *in vivo* would be very useful for investigating Arresten as a cancer therapeutic, and also for investigating Arresten as a biomarker for cancer detection.

2. MATERIALS AND METHODS

2.1 Cell Lines

H1299 (NCI-H1299, ATCC CRL-5803) and HCT 116 wild-type (WT) (ACT CCL-247) and p53-/- were cultured in Dulbecco's Modified Eagle Medium (DMEM; Wisent) which was supplemented with 10% fetal bovine serum (FBS; Sigma Aldrich) and 1% gentamycin sulphate (Wisent) and were incubated at 37°C with 5% CO_2.

2.2 Human Plasma Samples

Human plasma samples from patients with NSCLC and control patients were purchased from the Respiratory Health Research Network Biobank in Quebec City. All participants in this study completed research ethics training via completion of the Course on Research Ethics. This study was performed in accordance with the approved McGill Ethics Protocol.

2.3 Infection of Cells with Ad-p53 or Ad-lacZ

H1299 cells were seeded at a density of $1.5x10^6$ cells in a 10 cm plate overnight. The cells were then infected with either adenovirus expressing p53 (Ad-p53) or lacZ (Ad-lacZ) in infection media consisting of DMEM. After 24 hours, the cells were washed twice with 1x PBS, and left in 3 mL of DMEM + 1% insulin-transferrin-selenium (ITS) (Wisent) overnight. The next morning, cells and media were harvested. Both media and scraped cells were centrifuged at 1200 RPM for 5 minutes to pellet any cells in the media, and to pellet the scraped cells. Supernatant of the media containing tube was then transferred to a new tube and any pelleted cells in the media containing tube were combined with scraped cell pellet. The cell pellets were washed once with 1x PBS, and both the cell pellets and conditioned media were flash frozen using dry ice and 100% ethanol.

2.4 Treatment of Cells with 5-FU or CPT

HCT 116 WT and HCT 116 p53-/- cells were seeded at a density of $2x10^6$ cells in 10 cm plates overnight. Both HCT 116 WT and HCT 116 p53-/- cells were treated with either 400 µM 5-flurocil (5-FU), 100 nM camptothecin (CPT), or dimethyl sulfoxide solution (DMSO) as a control in DMEM with 10% FBS and 1x gentamycin for 24 hours. After 24, the cells were washed twice with 1x PBS, and left in 4 mL of DMEM + 1% ITS overnight. The next morning, cells and media were harvested. Both media and scraped cells were centrifuged at 1200 RPM for 5 minutes to pellet any cells in the media, and to pellet the scraped cells. Supernatant of the media containing tube was then transferred to a new tube and any pelleted cells in the media containing tube were combined with scraped cell pellet. The cell pellets were washed once with 1x PBS, and both the cell pellets and conditioned media were flash frozen using dry ice and 100% ethanol.

2.5 Treatment of Cells with GM6001 or Marimastat

H1299 cells were seeded at a density of $1.5x10^6$ cells in a 10 cm plate overnight. The cells were then infected with either Ad-p53 or Ad-lacZ. After twenty-four hours, the cells were washed twice with 1x PBS, and left in 3 mL of DMEM + 1% ITS + 25 µM GM6001 (Sigma-Aldrich) or 100 µM Marimastat (Sigma-Aldrich) overnight. The next morning, cells and media were harvested. Both media and scraped cells were centrifuged at 1200 RPM for 5 minutes to pellet any cells in the media, and to pellet the scraped cells. Supernatant of the media containing tube was then transferred to a new tube and any pelleted cells in the media containing tube were combined with scraped cell pellet. The cell pellets were washed once with 1x PBS, and both the cell pellets and conditioned media were flash frozen using dry ice and 100% ethanol.

2.6 Western Blot

Harvested cells were extracted with 1x buffer X (20mM Tris pH 7.4, 150 mM NaCl, 1% Triton-X-100) to obtain cell extracts. The concentrations of the soluble proteins in the cell extracts were quantified using the Pierce™ 660nm Protein Assay Reagent (ThermoFisher) and spectroscopy at 660 nm (VarioSkan). Equal concentrations of protein extracts were separated via sodium dodecyl sulphate-polyacrylamide gel electrophoresis, and proteins were transferred overnight at 30V to nitrocellulose membranes (Bio-Rad). Membranes were blocked in 5% milk in tris(hydroxymethyl)aminomethane-buffered saline in 0.5% Tween 20 (TBS-T). Membranes were then incubated in the appropriate primary antibody for the protein of interest (Table 1), and appropriate secondary antibodies. All primary and secondary antibodies were diluted in 5% milk in TBS-T. Proteins were detected using Clarity enhanced chemiluminescence (ECL) substrates (Bio-Rad) or Clarity Max Western ECL substrate (Bio-Rad). and imaged using the ChemiDoc XRS+ system (Bio-Rad) and then visualized using the ImageLab 6.0.1 Software (Bio-Rad).

Proteins in the conditioned media were precipitated at a ratio of 1:9 ratio of conditioned media to 100% ethanol, overnight at -20°C. The conditioned media was then centrifuged at 20 000 RCF for 15 minutes at 4°C and the supernatant was removed. Remaining proteins were resuspended in 1x Laemmli buffer and boiled for five minutes. Proteins were quantified using the Pierce™ 660nm Protein Assay Regent (Bio-Rad) and the Ionic Detergent Compatibility Reagent for the Pierce 660™ nm Protein Assay Reagent (Bio-Rad) and spectroscopy at 660nm (VarioSkan). Equal concentrations of protein were separated via TGX Stain-Free Fast Cast Acrylamide gel (Bio-Rad). Gels were activated using the ChemiDoc XRS+ system (Bio-Rad). Proteins were then transferred to low-fluorescence PVDF membranes using the Trans-Blot Turbo Transfer System (Bio-Rad). Membranes were blocked in 0.5% milk in TBS-T, incubated

in the appropriate primary antibody (Table 1) diluted in 0.5% milk in TBS-T, and incubated in the appropriate secondary antibody diluted in 0.5% milk in TBS-T. Proteins were detected using Clarity ECL substrates (Bio-Rad) or Clarity Max Western ECL substrate (Bio-Rad). and imaged using the ChemiDoc XRS+ system (Bio-Rad) and then visualized using the ImageLab 6.0.1 Software (Bio-Rad).

Table 1. Antibodies used for western blot.

Protein of Interest	Primary Antibody	Primary Antibody Supplier	Primary Antibody Dilution	Secondary Antibody	Secondary Antibody Dilution	MW on gel (kD)
p53	Mouse anti-p53 antibody (DO1): sc-126	Santa Cruz	1:1000	Anti-mouse HRP	1:10 000	53
Actin	Rabbit anti-actin	Sigma-Aldrich	1:10 000	Anti-Rabbit HRP	1:5000	42
Arresten	Monoclonal rat anti-COL4A1 (H11)	Sado, et al. (1995) [79]	1:30 000	Anti-Rat HRP	1:5000	80

2.7 Human ELISA

Human plasma samples from patients NSCLC and patients without cancer control samples, conditioned media from Ad-p53 and Ad-lacZ infected H1299 cells, and conditioned media from HCT 116 WT and p53-/- cells treated with 5-FU, CPT, or DMSO, were incubated at 37°C for two hours in wells of a commercial ELISA kit **(**MyBioSource; MBS903601**)**. The wells were then emptied and incubated with the kit's primary antibody, anti-biotin, for one hour at 37°C. The wells were then emptied and washed three times, using the kit's washing buffer, before being incubated with anti-horseradish peroxidase (HRP) antibody, the kit's secondary antibody, for one hour at 37°C. The wells were then emptied and washed five times. HRP substrate,

provided by the ELISA kit, was then added to the wells and the wells were incubated at 37°C for 15-30 minutes until appropriate colour development was observed. The kit's stop solution was added to the wells and the plate was read at optical densities (OD) of 450 nm and 540 nm (VarioSkan).

2.8 Statistical Analysis for ELISA

All graphs were created using Excel 2015. Arresten concentrations for conditioned media samples are presented as mean ± standard error of mean (SEM), with n=6. Arresten concentrations for all human plasma samples are presented in a dot plot. Statistical significance was calculated using two-tailed student's t-test.

2.9 Strategy for Creating *COL4A1* Conditional Knockout

Transcribing the *COL4A1* gene in mice can produce eight total transcripts, four of which produce functional proteins. However, there is no one region of the gene that is present in all four coding transcripts, making it impossible to delete only one region of the coding portion of the gene and have no functional protein. Based on the *COL4A1*/*COL4A2* knockout study in mice by Pöschl et al. (2004), it was determined that a *COL4A1* conditional knockout mouse model in C57BL/6J could be created by floxing out exon one of *COL4A1*, the promoter, and exon one of *COL4A2* (Figure 1). Ideally, only exon one of *COL4A1* and the promoter would have been floxed. However, due to the guideline for creating conditional knockout mice specifying that loxP sequences should not be inserted into an exon of a gene, it was determined that the first exon of *COL4A2* should be floxed too. Furthermore, it was determined that conditionally deleting the first exon of *COL4A2* would not be an issue due to two main reasons. First, by deleting the promoter of *COL4A1/2, COL4A2* could not be transcribed. Secondly, COL4A2

functions only in a heterotrimer with COL4A1. Thus, if *COL4A1* is knocked out, COL4A2 would not be able to form its only heterotrimer. To confirm the exact sequence of the critical exon in C57BL/6J mice, we performed PCR using Q5® High-Fidelity DNA Polymerase (NEB) with the following reaction components: 1x GC enhancer, 1x Q5 reaction buffer, 0.5 µM forward primer, 0.5 µM reverse primer, 200 µM dNTPs, 2 µl blastocyst DNA for round 1 and 1 µl PCR product for round 2, and 0.02 U/µl Q5 polymerase, to amplify the critical exon region and the region surrounding the critical exon. Primers and PCR conditions used for this PCR can be found in table 2. PCR products were then sequenced via Sanger sequencing using four primers (Table 2).

Table 2. PCR conditions used to amplify and sequence critical exon

PCR primers	**PCR Reaction Conditions**	**Sequencing Primers (5'-3')**
Forward: **CAGACGCTTTGAACCGCAC** **Reverse:** **TGTAGGACTCTGAGTGGCG**	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 63.1ºC for 10 sec 4. 72ºC for 20 sec 5. Repeat 2-4 for 34x 6. 72ºC for 10 min 7. 4ºC hold	1) TACAGGTCCGCGACTTGG 2) CAAATTGGGGAGCTCCTCG 3) ACGGCCAGGTGCATTCTTC 4) CGGGCCAACGCTTCTTCA

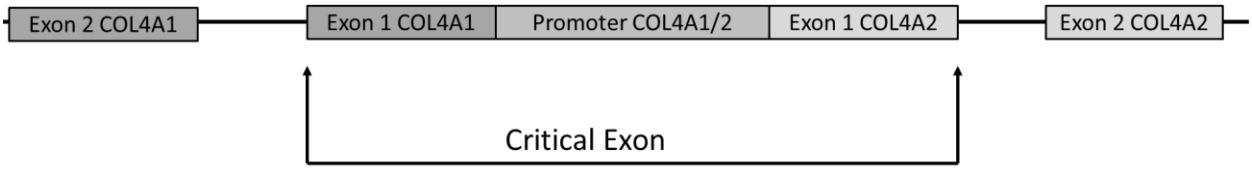

Figure 1. Diagram of critical exon for *COL4A1* floxed C57BL/6J strain.

2.10 sgRNAs and Repair Template Designs

Several factors were considered when designing the sgRNAs used for the generation of the CRISPR-edited floxed mouse strain: 1) design sgRNAs at least 100 base pairs away from any

exon/intron boundaries within the intron to reduce possibility of disruption of splice junction sites; 2) consider predicted efficiency and potential off-target effects of each sgRNA. The online sgRNA design web tool, CRISPOR.org, was used to aide in the sgRNA design process, as this program allowed for the several factors to be considered when designing potential sgRNAs that could be used to introduce the loxP sequences flanking the critical exon C57BL/6J mice [80].

The repair template used to produce the *COL4A1* 3'loxP strain was manually designed according to the following guidelines: 1) the loxP sequences were inserted at the Cas9 cut site to prevent Cas9 from cutting the DNA again when the double stranded cut is repaired with the repair template; 2) both loxP sequences were designed to be in the same orientation to ensure an excision, not inversion, of the critical exon; 3) both homology arms were designed to be approximately 200 base pairs each. This repair template was produced as a double stranded repair template cloned into the pUC57 plasmid via EcoRV blunt-end cloning (GenScript) (Figure 2A).

To insert the 5'loxP sequence into zygotes produced from IVF from the sperm of *COL4A1* 3'loxP mice and oocytes from *COL4A1* 3'loxP mice or WT mice, two different strategies were created to account for the "A" insertion mutation observed in the 3'loxP positive F1 mice. Strategy one used the previously tested, but not chosen, sgRNA "guide 2 upstream" which was 86 base pairs upstream from the previously used sgRNA. Strategy two used a new, untested sgRNA which had the same sequence as "guide 3 upstream", however it also included the "A" insertion mutation. Two single-stranded oligo donor (ssODN) repair templates (GenScript) were designed used to produce the floxed *COL4A1* strategy one and strategy two strains. These strains were developed were manually designed according to the following guidelines: 1) for both

ssODNs, the loxP sequence was inserted at the Cas9 cut site of the specific sgRNA for that strategy; 2) both ssODNs were designed to correct "A" insertion mutation (Figure 2B).

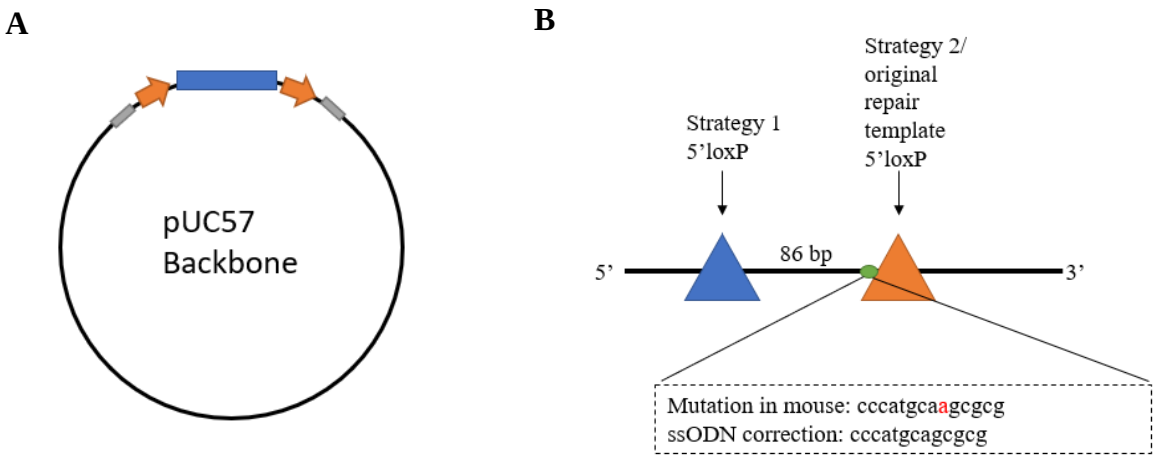

Figure 2. Diagrams illustrating the three repair templates used. (A) Original double stranded repair template used to produce the *COL4A1* 3'loxp strain was cloned into pUC57 vector. Grey boxes represent homology arms of 200 base pairs, orange arrows represent loxP sequences in the same orientation, and the blue box represents the critical exon. **(B)** Both ssODNs used to produce the floxed *COL4A1* containing strain. Both ssODNs were designed to correct "A" insertion mutation, indicated by the red letter in the "mutation in mouse" sequence, observed in *COL4A1* 3'loxP strain. However, the ssODN for strategy one has the 5'loxP sequence 86 base pairs upstream from the ssODN for strategy two and the original double stranded repair template.

2.11 Animals

All CRISPR-edited mice were produced in the C57BL/6J strain purchased from The Jackson Laboratory and were created and initially housed in the McGill Transgenic Core Facility. Founders were then transferred to room 307E of the Goodman Cancer Research Centre Mouse Facility where any matings of mice occurred, and any offspring were housed. All animal studies were performed in accordance to approved protocols.

2.12 Microinjections and Electroporation

To test the efficiency of each guide, the McGill Transgenic Core performed pronuclear injections to inject Cas9 recombinant protein and an *in vitro*–transcribed sgRNA into C57BL/6J zygotes at the two-cell stage which were then cultured to the blastocyst stage. At the blastocyst stage DNA was extracted from the blastocyst. The extracted DNA was subjected to nested PCR using the Q5® High-Fidelity DNA Polymerase (NEB) with the following reaction components: 1x GC enhancer, 1x Q5 reaction buffer, 0.5 µM forward primer, 0.5 µM reverse primer, 200 µM dNTPs, 2 µl blastocyst DNA for round 1 and 1 µl PCR product for round 2, and 0.02 U/µl Q5 polymerase. These PCRs were used to amplify the region surrounding where DNA would have been cut (Tables 3 and 4). The amplified region was then sequenced via Sanger sequencing using the round 2 forward primer to determine if there were any cuts in the DNA in the appropriate region.

To generate the *COL4A1* 3'loxP founder mice, C57BL/6J zygotes at the two-cell stage were microinjected with two pre-assembled ctRNP complexes which contain one sgRNA that flanks the critical exon, the Cas9 protein, and the repair template by the McGill Transgenic Core. Mouse embryos were then implanted into pseudo pregnant C57BL/6J females.

To generate the floxed *COL4A1* mice, sperm was harvested from an N1F1 male mouse containing the 3'loxP insert. This sperm was used to fertilize oocytes from N1F1 and N1F2 females containing the 3'loxP insert and oocytes from WT females. At the one-cell stage, either microinjection or electroporation were used to insert the ssODN and a pre-assembled ctRNP complex containing the sgRNA and Cas9 protein. Embryos were then transferred to pseudo pregnant C57BL/6J females.

Table 3. Round 1 of nested PCR conditions for screening of blastocysts.

PCR Reaction	Primers (5'-3')	Conditions
Upstream sgRNAs 1,3	Forward: GAAAGCGACCAAAACATGCGG (Sigma-Aldrich) Reverse: CTTGGTGTCGCAGGGAGTCC (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 64.4ºC for 10 sec 4. 72ºC for 20 sec 5. Repeat steps 2-4 for 34X 6. 72ºC for 10 min 7. 4ºC hold
Upstream sgRNA 2	Forward: CTTACCCCCAAGACGCTCTG (Sigma-Aldrich) Reverse: AAGCACCAGCAAGGCACGC (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 69.2ºC for 10 sec 4. 72ºC for 20 sec 5. Repeat steps 2-4 for 34X 6. 72ºC for 10 min 7. 4ºC hold
Downstream sgRNA 1,2,3	Forward: TGGCTTCTGCTGCTCTTCGC (Sigma-Aldrich) Reverse: AAGGCAAGAGTCGAGGCAG (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 64.0ºC for 10 sec 4. 72ºC for 20 sec 5. Repeat steps 2-4 for 34X 6. 72ºC for 10 min 7. 4ºC hold

Table 4. Round 2 of nested PCR conditions for screening of blastocysts.

PCR Reaction	Round 2 Primers (5'-3')	Round 2 Conditions
Upstream sgRNAs 1,3	Forward: CAGACGCTTTGAACCGCAC (IDT) Reverse: AAGCACCAGCAAGGCACGC (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 71.1ºC for 10 sec 4. 72ºC for 20 sec 5. repeat steps 2-4 for 34X 6. 72ºC for 10 min 7. 4ºC hold
Upstream sgRNA 2	Forward: GAAAGCGACCAAAACATGCGG (Sigma-Aldrich) Reverse: CGGGCCAACGCTTCTTCAG (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 64.4ºC for 10 sec 4. 72ºC for 20 sec 5. Repeat steps 2-4 for 34X 6. 72ºC for 10 min 7. 4ºC hold
Downstream sgRNA 1,2,3	Forward: GCAGCTGCGAAGGTGAGTTC (Sigma-Aldrich) Reverse: GTTCTCAAGCGTGCCGTAG (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 65.3°C for 10 sec 4. 72ºC for 20 sec 5. Repeat steps 2-4 for 34X 6. 72ºC for 10 min 7. 4ºC hold

2.13 Mouse DNA Extractions

For guide testing, DNA was extracted from blastocysts by the Transgenic Core. To screen founder mice and F1 generation mice, ear or tail tissue was extracted with Proteinase K (30 µg) in a tissue lysis buffer (100mM Tris-HCl, pH 8.0, 5mM EDTA, 0.2% SDS, 200mM NaCl) overnight at 55°C. The next day, the Proteinase K was heat inactivated via boiling of the samples at 95°C for 5 minutes. The samples were then diluted with sterile water (Wisent) to a total volume of 800 µl.

2.14 Founder and F1 Generation Screening

The screening strategy for all founder mice and any F1 generation mice involved the use of three different PCRs and sequencing the products of these PCR via Sanger sequencing. The first PCR, termed the 5'loxP PCR, amplified the region surrounding the 5'loxP insert. The second PCR, termed the 3'loxP PCR, amplified the region surrounding the 3'loxP insert. Both the first and second PCRs use one primer within the repair template and one outside the repair template, thus providing confirmation of insertion of the loxP sequences in the desired location. A third PCR, termed the loxP-loxP PCR, amplified the critical exon region. Two different loxP-loxP PCRs were performed. LoxP-loxP PCR #1 amplified the critical region for the *COL4A1* 3'loxP mice and the floxed *COL4A1* mice generated by strategy two. LoxP-loxP PCR #2 amplified the critical exon region for the floxed *COL4A1* mice generated by strategy one. This PCR uses primers that overlap part of a homology arm and part of the repair template to confirm that both loxP sequences are on the same allele (Figure 3). PCR reactions were performed using the Q5® High-Fidelity DNA Polymerase (NEB) with the following reaction components: 1x GC enhancer, 1x Q5 reaction buffer, 0.5 µM forward primer, 0.5 µM reverse primer, 200 µM dNTPs, 3 µl diluted DNA from digested mouse tissue, and 0.02 U/µl Q5 polymerase. PCR conditions and primers specific to each reaction can be found in Table 4.

Table 5. PCR conditions for founder and f1 generation screening PCRs.

PCR Reaction	Primers Used (5'-3')	Reaction Conditions
5'loxP PCR	Forward: GAAAGCGACCAAAACATGCGG (Sigma-Aldrich) Reverse: CGGGCCAACGCTTCTTCAG (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 71.1ºC for 10 sec 4. 72ºC for 20 sec 5. repeat 2-4 34x 6. 72º for 10 min 7. 12º hold
3'loxP PCR	Forward: TGGCTTCTGCTGCTCTTCGC (Sigma-Aldrich) Reverse AAGGCAAGAGTCGAGGCAG (Sigma-Aldrich)	1. 98ºC for 30 sec 2. 98ºC for 5 sec 3. 64ºC for 10 sec 4. 72ºC for 20 sec 5. Repeat 2-4 34x 6. 72º for 10 min 7. 12º hold
loxP-loxP PCR #1	Forward: CTCAGCCCATGCATAACTTC (IDT) Reverse: GAAGTTATGCGGGGTGAGGGT (IDT)	1. 98°C for 30 sec 2. 98°C for 5 sec 3. 65°C for 10 sec 4. 72°C for 20 sec 5. Repeat 2-4 for 34x 6. 72°C for 10 min 7. 4°C hold
loxP-loxP PCR #2 *Did not work	Forward: CGTTTCCTTGCATAACTTCG (Sigma-Aldrich) Reverse: GAAGTTATGCGGGGTGAGGGT (IDT)	1. 98°C for 30 sec 2. 98°C for 5 sec 3. 56-71°C for 10 sec 4. 72°C for 20 sec 5. Repeat 2-4 for 34x 6. 72°C for 10 min 7. 4°C hold

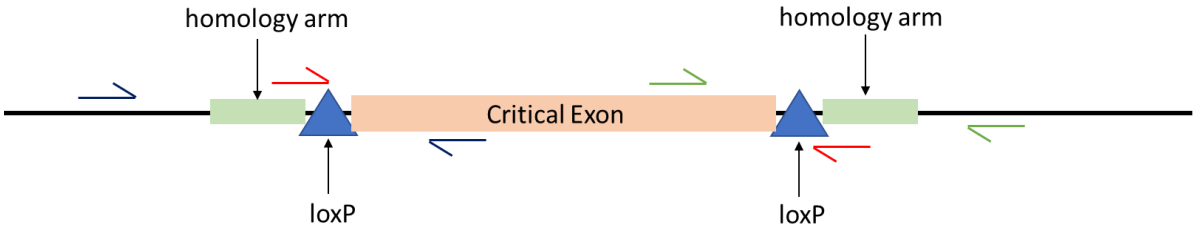

Figure 3. Schematic for screening strategy of founder and F1 mice. The blue set of primers illustrate the PCR used for the 5'loxP PCR. The green set of primers illustrate the PCR used for the 3'loxP PCR. The red set of primers illustrate the PCR used for the loxP-loxP PCR.

2.15 Mouse Matings

WT female mice were mated with the F0 *COL4A1* 3'loxP mouse 5478 to produce the N1F1 generations. Mice from the N1F1 strain that carried the 3'loxP insert were mated with WT mice to produce the N2F1 generation of mice.

2.16 *In-vitro* Cre Assay

The McGill Transgenic Core performed an *in-vitro* Cre assay on genomic founder mouse DNA. Genomic founder mouse DNA was digested with 2 units of Cre recombinase (NEB) overnight at 37°C. PCR was then used to amplify the region that would contain the loxP sequences flanking the critical exon for digested and nondigested products (Table 6). These samples were then resolved by electrophoresis on a 1.5% agarose gel and sequenced via Sanger sequencing.

Table 6. Primers used for *in-vitro* Cre assay PCR.

Primer Name	Primer Sequence (5'-3')
Primer cre FW 1	TAGCACCCTGACTCCGCATA
Primer cre FW 2	CAAAGCGGGTTAGCAAAGGT
Primer cre RV 1	AGAAACACACCCAAAGCGGT
Primer cre RV 2	ACCATCACGGTCTCCAAACA

3. RESULTS

3.1 Arresten as a biomarker for cancer detection

3.1.1 Inducing expression of p53 *in vitro* results in an increase in secreted Arresten

In order to ascertain the relationship between p53 and Arresten, the expression of p53 was induced in two cell lines to determine the effect on secreted Arresten production. The first cell line used was the H1299 NSCLC cell line. This cell line is a lung cancer cell line which has a homozygous partial deletion of the p53 gene, and thus does not express the p53 protein. These cells were infected with Ad-p53, to induce p53 expression in these cells, and as a control for infection, H1299 cells were also infected with Ad-lacZ. The cells were then analyzed via western blot to confirm p53 expression in the cells that were infected with Ad-p53 and lack of p53 expression in the cells infected with Ad-lacZ (Figure 4A). The conditioned media of the cells was also analyzed via ELISA to determine the effect of inducing p53 expression on secreted Arresten levels. The results of the ELISA show that induction of p53 results in a concentration of 56.17 ng/ml of secreted Arresten whereas the concentration of secreted Arresten in the control cells that do not express p53 is 0.31 ng/ml. This difference represents a dramatic increase in in secreted Arresten levels in cells that express p53 versus cells that do not express p53. (Figure 4B).

The second cell line used was the HCT 116 colorectal carcinoma cell line. This cell line is also a human cancer cell line; however, the WT version of these cells can express p53, whereas the p53-null cell line has been genetically modified to produce homozygous disruption of the p53 gene, resulting in an inability to express the p53 protein [81]. Both the WT and p53-null cell lines were treated with either 5-FU or CPT to endogenously induce p53 expression to confirm

that the increase in secreted Arresten due to p53 expression could be observed in an endogenous setting, or DMSO as a control. 5-FU stabilizes p53 and induces p53-dependent genes**,** whereas CPT induces p53 expression by inhibiting DNA topoisomerase I [82]. These cells were also analyzed by western blot to confirm the expression of p53 for WT cells treated with 5-FU or CPT, and the lack of p53 expression of WT control cells, and all p53-null cells (Figure 4C). The conditioned media of these cells was analyzed via ELISA to determine the effect of endogenously expressed p53 on secreted Arresten levels. The results show that WT and p53-null control cells have similar levels of secreted Arresten, concentrations of 0.42 ng/ml and 0.47 ng/ml respectively, where this difference is not statistically significant. However, for the WT and p53-null cells treated with 5-FU, the secreted Arresten concentrations are 5.79 ng/ml and 0.97 ng/ml respectively which is a statistically significant difference. Furthermore, for the WT and p53-null cells treated with CPT, the secreted Arresten concentration are 2.97 ng/ml and 1.17 ng/ml respectively, which is also a statistically significant difference. Lastly, the secreted Arresten concentrations of the p53 null cells treated with both 5-FU and CPT are both not statistically significant from both the WT and p53-null control secreted Arresten concentrations (Figure 4D).

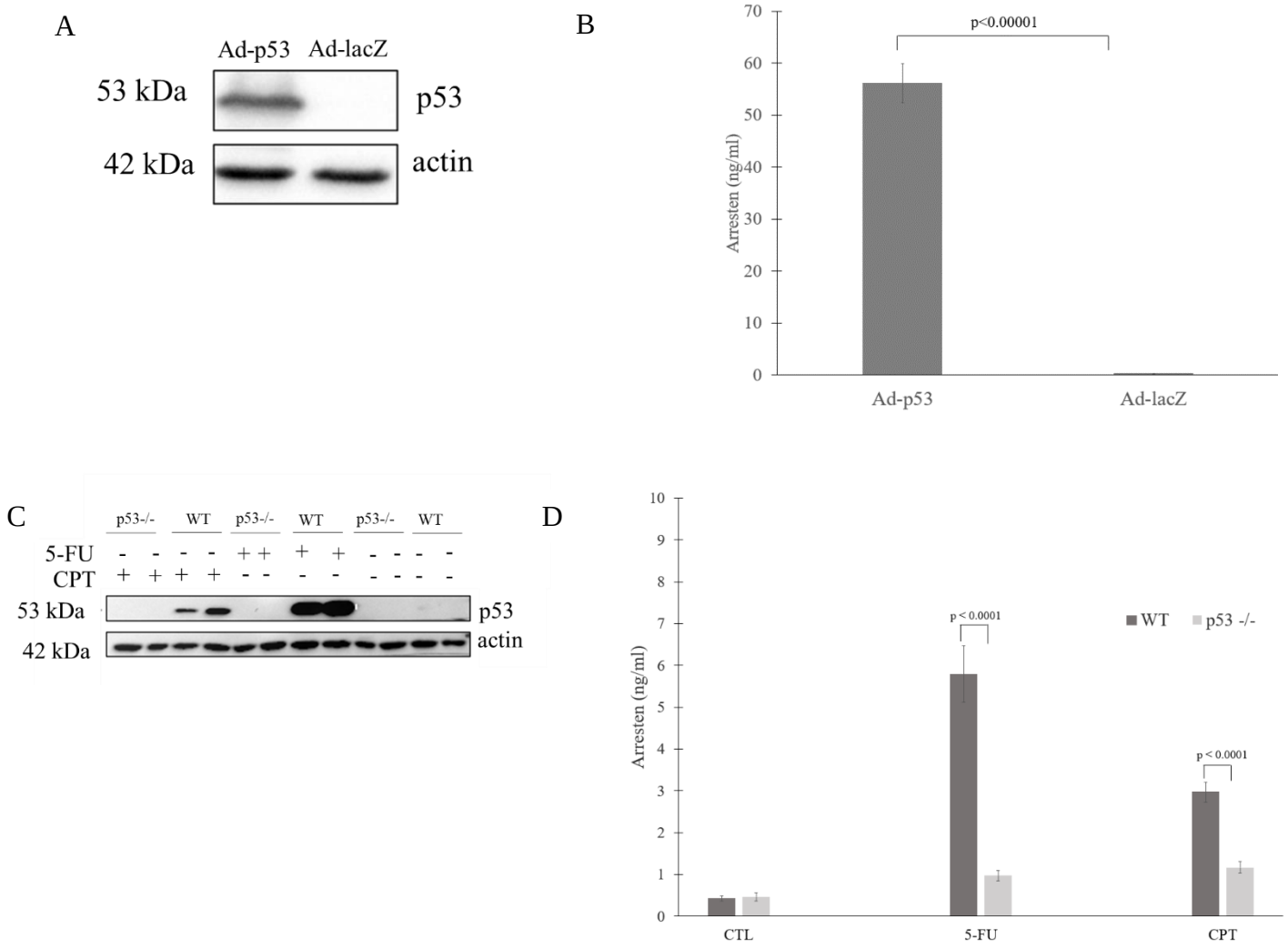

Figure 4. Secreted Arresten quantification by ELISA after induction of p53 expression. (A) Western blot analysis for p53 expression in H1299 cells infected with Ad-p53 or Ad-lacZ. **(B)** ELISA results for concentrations (ng/ml) of Arresten in the conditioned media from H1299 cells infected with Ad-p53 or Ad-lacZ. **(C)** Western blot analysis for p53 expression in HCT 116 WT and p53-/- cells treated with 5-FU, CPT or DMSO (control). **(D)** ELISA results for concentrations (ng/ml) of Arresten in the conditioned media from HCT 116 WT and p53-/- cells treated with 5-FU, CPT, or DMSO (control). Data is presented as mean ± SEM, where n=6, and p values are calculated using Student's t-test for comparison between samples where p53 is induced and where p53 is not induced.

3.1.2. Patients with adenocarcinoma NSCLC may have lower plasma Arresten levels compared to control patients

Previous unpublished work in our lab has suggested that patients with NSCLC may have lower Arresten levels in their plasma compared to patients with no cancer, thus, suggesting Arresten may have some merit as a biomarker for cancer detection. However, this study was a preliminary study, and no information was given regarding the histological subtype of NSCLC, or any other patient variables. Thus, it was uncertain if this relationship was dependent on the histological subtype of NSCLC, or any other variables. In order to further investigate this potential relationship of lower Arresten levels in plasma of patients NSCLC and Arresten's potential as a biomarker for cancer diagnosis, plasma samples from patients with NSCLC taken at first diagnosis of NSCLC and control plasma samples from patients with no cancer were obtained from the Respiratory Health Research Network Biobank in Quebec City for ELISA**.** All samples obtained had additional accompanying information such as histological subtype, biological sex of patient, smoker status, histological grade of cancer, etc. and thus these samples were more highly controlled allowing for true investigation into this potential relationship. The results of ELISA show that the mean Arresten concentrations for the different histological subtypes of NSCLC and the control patients are as follows: 11.94 ng/ml for adenocarcinoma, 15.49 ng/ml for large cell carcinoma, 20.50 ng/ml for squamous cell carcinoma, 17.97 ng/ml for small cell carcinoma, and 15.35 ng/ml for control patients with no cancer, in which there is no statistically significant difference between control patients and any histological subtype of NSCLC at $p<0.05$ (Figure 5). However, Arresten concentrations in plasma samples from adenocarcinoma patients are statistically lower compared to control patients at $p<0.1$.

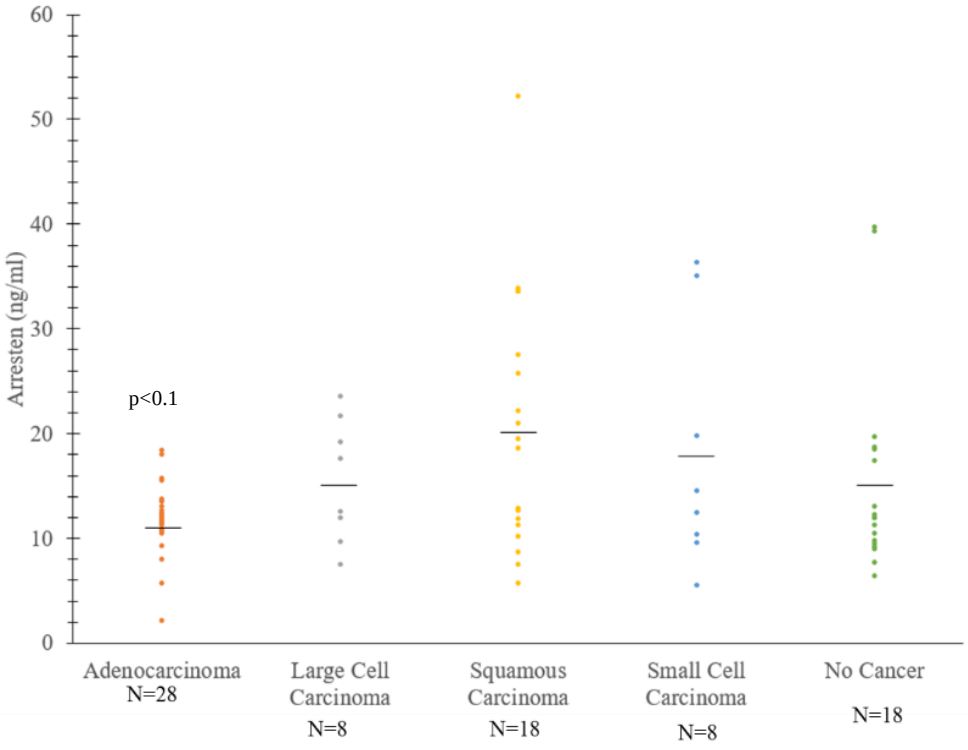

Figure 5. Comparison of Arresten concentration in plasma of NSCLC patients to plasma of patients with no cancer. Arresten concentrations in plasma of NSCLC patients by differing NSCLC histological subtype and control patients as measured by ELISA. Data presented in dot plot with means of each subgroup. Student's two-tailed t-test used to determine significance.

3.2 Arresten may be processed from COL4A1 by a secreted protease inhibited by GM6001

The processing of Arresten is currently unknown; however, it is known that MMPs play an important role in the processing of other collagen derived matrikines. Therefore, in order to investigate how Arresten is processed in the ECM, two broad-spectrum MMP inhibitors, GM6001 and marimastat were tested [83]. Both MMP inhibitors have been shown to target some of the same secreted proteases, not limited to MMPs, and some different secreted proteases. To test the inhibitors, H1299 cells were infected with either Ad-p53 to induce p53, and therefore Arresten, expression or Ad-lacZ as a control, and then treated with either GM6001 or marimastat 24 hours later. Western blots were done for both experiments to confirm the expression of p53 in the cells infected with Ad-p53 and the lack of p53 expression for the cells infected with Ad-lac (Figures 6A, 6B). Furthermore, western blots were done to determine the how treatment of the cells with either MMP inhibitor affects the level of Arresten secreted in into the conditioned media. Western blot results for proteins in the conditioned media show that when H1299 cells infected with Ad-p53 are treated with GM6001, the cells secreted a lower level of Arresten compared to cells infected with Ad-p53 but are not treated with GM6001 (Figure 6C). Conversely, for H1299 cells infected with Ad-p53 and treated with marimastat, western blot results for proteins in the conditioned media show that these cells secreted a similar level of Arresten as H1299 cells infected with Ad-p53 but not treated with marimastat (Figure 6D).

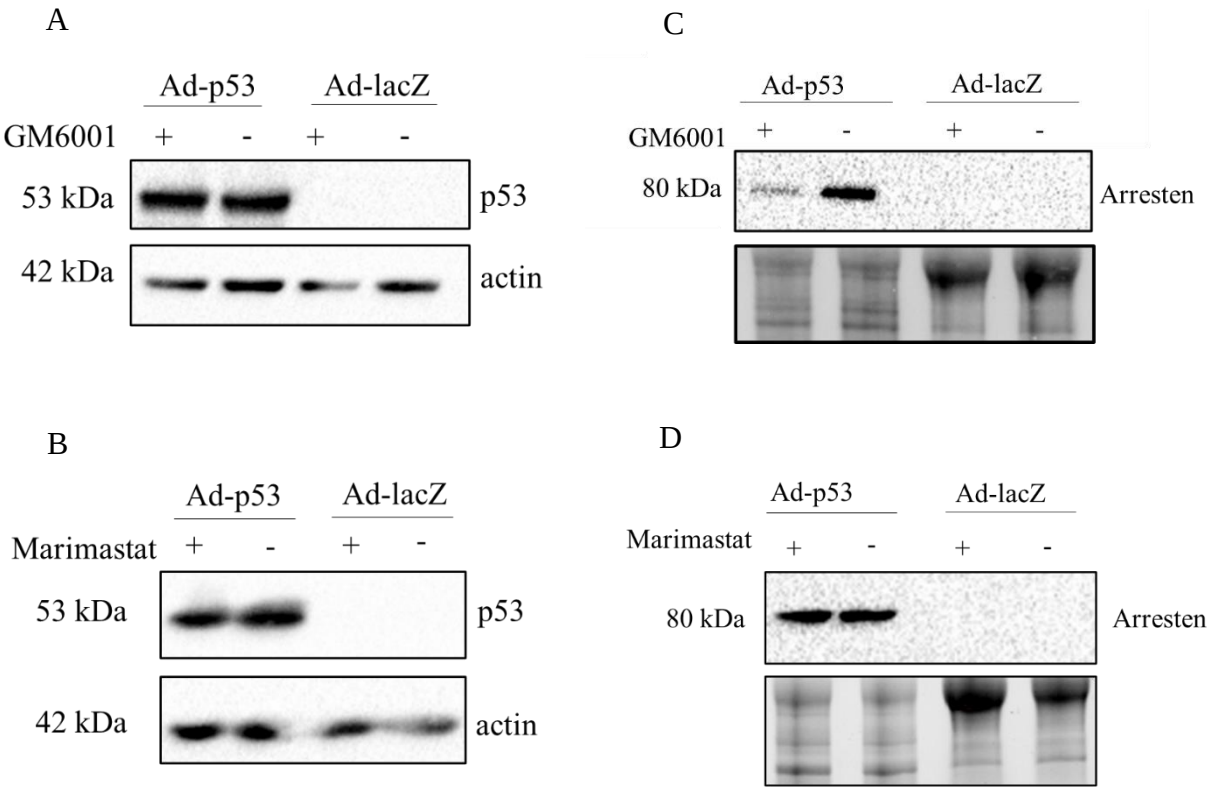

Figure 6. Treatment of H1299 cells infected with Ad-p53 or Ad-lacZ with GM6001 or marimastat. (A) Western blot analysis for p53 expression in H1299 cells infected with Ad-p53 or Ad-lacZ and treated with GM6001. **(B)** Western blot analysis for p53 expression in H1299 cells infected with Ad-p53 or Ad-lacZ and treated with marimastat. **(C)** Western blot results for Arresten in the conditioned media of H1299 cells infected with Ad-p53 or Ad-lacZ and treated with GM6001. Total protein indicates equal loading of protein for all samples, and n=3. **(D)** Western blot results for Arresten in the conditioned media of H1299 cells infected with Ad-p53 or Ad-lacZ and treated with marimastat. Total protein indicates equal loading of protein for all samples, and n=3.

3.3 Creating a floxed *COL4A1* mouse strain

3.3.1 Testing sgRNAs in blastocysts allows for more accurate selection of sgRNAs for generation of floxed mice

In order to introduce the loxP sequences into a mouse genome, CRISPR-Cas9 presents a novel way to create a conditional knockout mouse compared to traditional methods, such as using ES cells. Using CRISPR-Cas9 directed genome editing to generate floxed alleles is achieved by injecting mouse zygotes with two pre-assembled complexes called ctRNP complex, in which each complex contains one sgRNA and the Cas9 protein, and a repair template. Each sgRNA must be designed so that it targets one region flanking the critical exon where the loxP sequences should be introduced.

To design the sgRNAs, the online CRISPR design program CRIPSOR.org developed by Concordet and Haeussler (2018) was used [80]. This online program works by determining potential sgRNAs from an input sequence, and then ranking the sgRNAs with different scores corresponding to predicted efficiency, predicted specificity, and off-target effects. For both the predicted efficiency and predicted specificity scores, a higher score corresponds to higher predicted efficiency and higher predicted specificity. After comparing the scores of all presented sgRNAs that could be used to introduce a 5'loxP sequence and a 3'loxP sequence, three 5' targeting sgRNAs and three 3' targeting sgRNAs were chosen to test for guide efficiency (Table 7).

Guide efficiency was tested by using pronuclear injections to inject recombinant Cas9 protein and *in vitro* transcribed gRNA into C57BL/6J zygotes at the two-cells stage, and then culturing the zygotes to the blastocyst stage, at which point genomic DNA was extracted. Using

the extracted DNA, the region surrounding the area where the DNA would have been cut was then amplified via PCR and sequenced via Sanger sequencing. In total 12 blastocysts for “guide 1 upstream”, 11 blastocysts for “guide 2 upstream”, 11 blastocysts for “guide 3 upstream”, 8 blastocysts for “guide 1 downstream”, 9 blastocysts for “guide 2 downstream”, and 11 blastocysts for “guide 3 downstream” were screened to determine if there were any indels or mutations present in the DNA around the cut site, which would indicate the success of each sgRNA. For each sgRNA, guide efficiency was then calculated by dividing the number of successfully cut blastocyst by the total number of blastocysts screened for that guide. The results of the sgRNA testing indicate that “guide 3” is the most efficient upstream or “5’loxP” guide, whereas for the downstream, or “3’loxP”, guides, “guide 1” and “guide 3” were equally efficient (Table 8). Thus, “guide 3 upstream” was chosen for the upstream sgRNA, and arbitrarily, “guide 1 downstream” was chosen for the downstream sgRNA. These results were contrary to which guides were predicted to be the most efficient by CRISPOR.org. CRISPOR.org predicted that “guide 1 upstream” for the upstream guides would have been the most efficient, whereas “guide 3 upstream’ would have been the least efficient. However, this was not observed, as in fact the results showed that “guide 3 upstream” was the most efficient, whereas “guide 1 upstream” was the least efficient.

Table 7. CRISPOR.org results for the three upstream sgRNAs and three downstream sgRNAs chosen to test for guide efficiency

gRNA ID	Guide sequence (+ PAM)	MIT specifi-city score	CFO specifi-city score	Predicted Efficiency		Outcome		Off targets for 0-1-2-3-4 mismatches (+ next to PAM)
				Doench'16	Mor-Mateos	Out-of-frame	Lindel	
gRNA up 1	GAATAGGGATCGCGCTGCAT GGG	97	99	45	74	63	65	0-0-0-4-19
gRNA up 2	AGAACTTTGGTCACCCCGCA AGG	93	95	55	48	71	84	0-0-0-5-49
gRNA up 3	GGAATAGGGATCGCGCTGCA TGG	95	96	41	35	64	88	0-0-0-2-69
gRNA down 1	GCTTAGCGTGTACCACGGCG GGG	98	97	65	61	71	79	0-0-0-0-19
gRNA down 2	TCGGTACCCTCACCCCGCCG TGG	94	97	64	63	49	65	0-0-0-3-24
gRNA down 3	AAGCTTAGCGTGTACCACGG CGG	95	97	68	84	60	77	0-0-0-3-27

Table 8. Results of sgRNA efficiency testing

gRNA ID	Number of cut embryos	% Efficiency
gRNA up1	2	16
gRNA up2	5	45
gRNA up3	8	73
gRNA dn1	8	100
gRNA dn2	8	89
gRNA dn3	11	100

3.3.2 Use of CRISPR-Cas9 allows for successful integration of both loxP sequences on the same allele

Generation of the floxed *COL4A1* strain was attempted by the McGill Transgenic Core via microinjection of C57BL/6J zygotes at the two-cell stage with two pre-assembled ctRNP complex, each containing either the upstream or downstream sgRNA, Cas9 protein, and the double-stranded repair template. From this round of injections, 12 different mice were screened using the screening strategy of the 5'loxP PCR, 3'loxP PCR, and loxP-loxP PCR. The first PCR performed was the 5'loxP PCR (Figure 7A). The results of this PCR show that unfortunately there were potentially only two mice positive for the upstream loxP insertion, and only one out of the two mice appeared to have a high frequency of insertion. The second PCR performed was the 3'loxP PCR (Figure 7B). The results of this PCR show that there were five mice likely positive for the downstream loxP insertion. However, there was only one mouse that was likely positive for the upstream loxP insertion and the downstream loxP insertion, mouse 5478, a male. Nonetheless, the loxP-loxP PCR was performed for all samples, and four different mice tested positive for this PCR, including mouse 5478 (Figure 7C). This was odd, as the PCR results for the 5'loxP PCR and 3'loxP PCR indicated that all mice, except for 5478, should not have been positive the loxP-loxP PCR.

Sanger sequencing was then used to sequence the region amplified by the 5'loxP PCRs and the 3'loxP PCR for all four mice (data not shown). The sequencing results revealed that mouse 5485 only had WT sequences for both regions, indicating that the positive result observed in the loxP-loxP PCR was likely due to an insertion of the repair template at a region that was not the target region. However, for mice 5414, 5478, and 5484 the sequencing results were inconclusive due to the mosaicism of these mice.

To obtain a conclusive analysis regarding mice 5414, 5478, and 5484, the Transgenic Core aided us in performing an *in vitro* Cre assay, where the Cre enzyme is added to digested genomic DNA isolated from each mouse, and then the region flanking the putative loxP sequences should be inserted is amplified by PCR. If both loxP sequences are present in the genome at the correct location, the addition of the Cre enzyme should result in recombination of the loxP sequences, and thus show a band on an agarose gel that is of smaller size than the sequence prior to digestion with the Cre enzyme. These results can then be further confirmed via Sanger sequencing. The results of this *in vitro* Cre assay showed a successful Cre digestion of the genomic DNA of mouse 5478 on agarose gel (Figure 7D) and via Sanger sequencing (data not shown). The results also showed mouse 5414 to be positive; however, we were skeptical of these results as the digested band was present in the "before digestion" sample too. Overall, based on our results from the three screening PCRs, the Sanger sequencing results, and the *in vitro* Cre assay, it was determined that mouse 5478 was the most likely candidate to be a founder mouse.

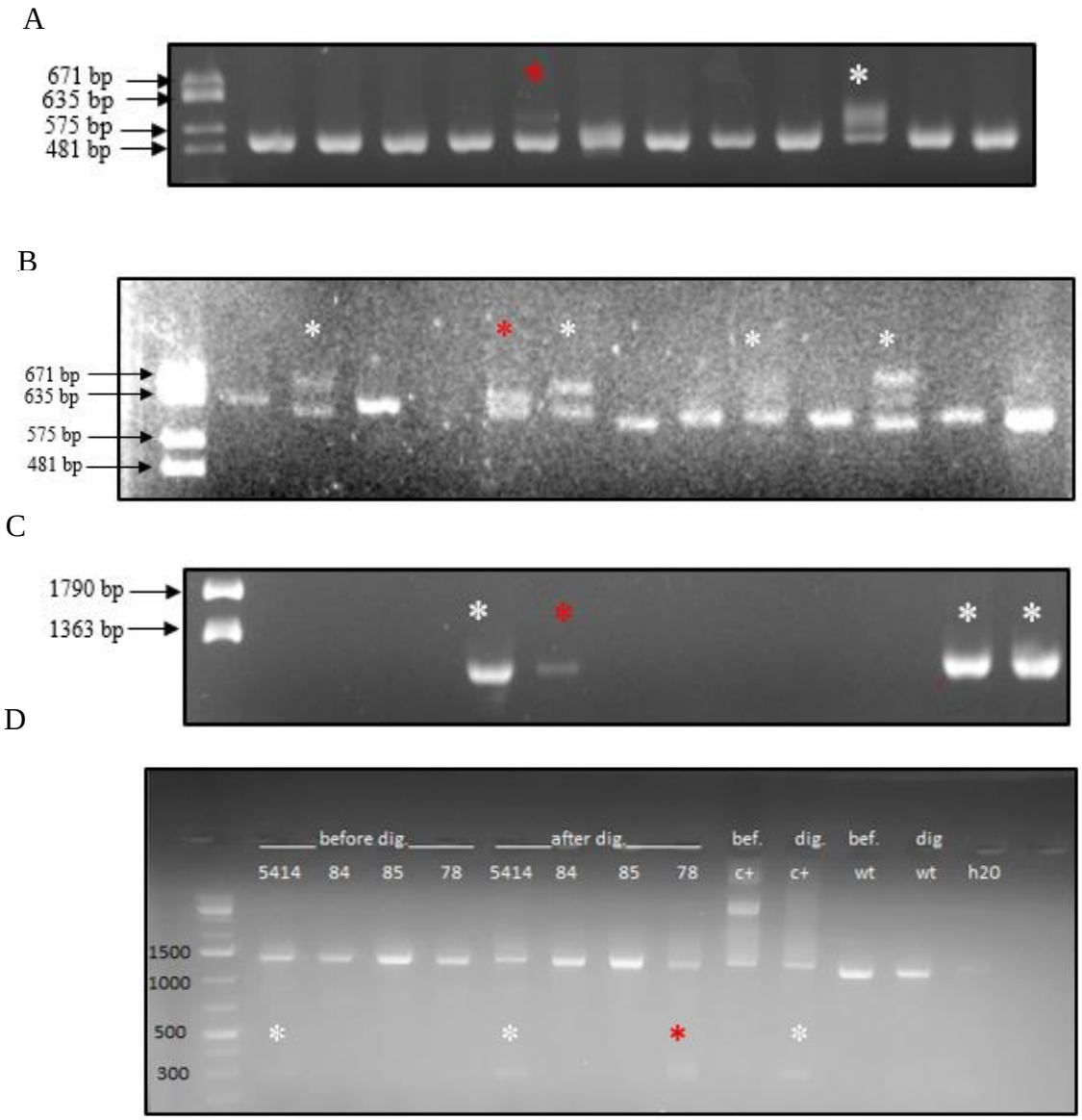

Figure 7. Screening of founder mice for first round of injections using double-stranded repair template. Positive mouse 5478 is indicated by red asterisk and other positive samples are indicated by a white asterisk. **(A)** 5'loxP PCR results resolved by electrophoresis on 3% agarose gel. **(B)** 3'loxP PCR results resolved by electrophoresis on 3% agarose gel. **(C)** loxP-loxP PCR results resolved by electrophoresis on 2% agarose gel. **(D)** *In-vitro* Cre assay results resolved by electrophoresis on 1.5% agarose gel.

3.3.3 F1 generation that does not have both loxP insertions passed down can be used to create the floxed strain

Mouse 5478 was mated to a WT C57BL/6J female mouse to produce the F1 generation. In total three F1N1 generations from this mating were produced. All three litters were screened using the three screening PCRs. Unfortunately, neither the 5' nor the 3' loxP insertions were passed down to litters two and litter three of the F1 generation (results not shown). However, for litter one, the 3'loxP PCR appeared to be positive for some of the mice, whereas the 5'loxP PCR was negative for all mice (Figures 8A, 8B).

After observing the lack of transmissibility of both the 5' and 3' loxP insertions to the F1 generation, it was concluded that mouse 5478 was most likely not going to produce any F1 litters where both the 5' and 3' loxP insertions were passed down. This was especially true when we again observed the 5'loxP result on agarose gel and saw that it appeared as though the 5'loxP was most likely not inserted at a high frequency in this mosaic mouse. As a result, it was decided to retire mouse 5478 as a breeder, and instead use sperm and oocytes from the *COL4A1* 3'loxP positive F1 generation for *in vitro* fertilization (IVF) and insertion of the 5'loxP sequence into these zygotes.

In order to use the 3'loxP positive F1 generation for IVF and insertion of the 5'loxP sequence into the IVF generated zygotes, the 5'loxP region was sequenced via Sanger sequencing to ensure there were no mutations. Surprisingly, sequencing results of the 5'loxP region revealed the mice were heterozygous for an "A" insertion mutation in a critical region. The mutation was within the region of where the upstream gRNA was supposed to anneal (Figure 8C). To try and determine the origin of this "A" insertion mutation, the 5'loxP region was sequenced in purchased WT C57BL/6J mice. Sanger sequencing revealed that these mice did not have the A insertion mutation observed in the 3'loxP positive F1 mice (data not shown).

Furthermore, the 5'loxP region was sequenced from mice produced from the initial round of injections but appeared to not have the 5'loxP sequence inserted when observing the agarose gel results. Sanger sequencing results of these mice showed that these mice did contain the "A" insertion mutation observed in the *COL4A1* 3'loxP positive F1 mice (data not shown).

The 3'loxP region also had to be sequenced via Sanger sequencing to ensure there were no mutations, and to confirm the orientation of the loxP sequence. Sequencing of the 3'loxP region was deemed to be inconclusive due to the difficulty in reading the chromatograms produced from Sanger sequencing. Therefore, an alternative PCR was performed to amplify only the loxP containing sequence and thus make it possible to draw conclusions about the sequence of this region. This alternative PCR used the forward primer from the 3'loxP PCR and the reverse primer from the loxP-loxP PCR. Sanger sequencing results of the region amplified by this PCR were successful in ensuring no mutations in this region and confirming the orientation of the loxP insertion (Figure 8D).

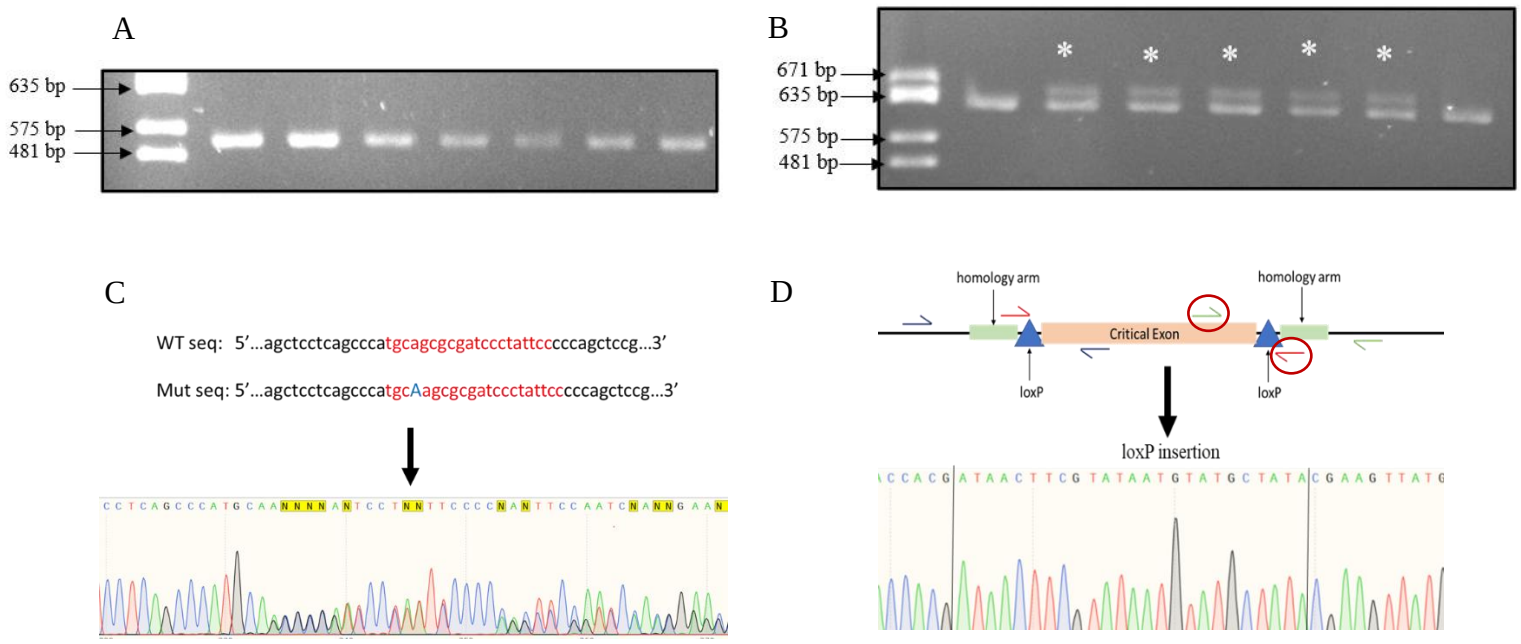

Figure 8. Screening of F1 generation litter one from founder mouse 5478. (A) 5'loxP PCR results resolved by electrophoresis on 3% agarose gel. **(B)** 3'loxP PCR results resolved by electrophoresis on 3% agarose gel. Positive samples are indicated by white asterisk. **(C)** Illustration of "A" insertion mutation. Red sequence for both the WT sequence and mutant sequence indicates the sequence where the upstream sgRNA is supposed to anneal, and the blue "A" in the mutant sequence indicates the "A" insertion mutation. Below is the chromatogram from the Sanger sequencing results, showing the heterozygous nature of the "A" insertion. **(D)** PCR performed to amplify only 3'loxP containing allele. Diagram shows the primers used for this PCR, indicated by the circled primers. Below is the chromatogram from the Sanger sequencing results showing the insertion of the downstream loxP sequence.

3.3.4 Generation of floxed *COL4A1* mice

In order to account for the "A" insertion mutation found in the region where the upstream sgRNA anneals in the 3'loxP positive F1 mice, two different strategies were developed to insert the 5'loxP sequence. The first strategy, termed "strategy one", involved using a previously tested sgRNA that was not chosen. This sgRNA is "guide 2 upstream" and it is 86 base pairs upstream of the old guide, "guide 3 upstream". Thus, a new ssODN was also used to insert the loxP at the correct location, and to try and correct the "A" insertion mutation. The second strategy, termed "strategy two", involved using a new, untested sgRNA. This sgRNA had the exact same sequence as "gRNA up 3", however it also included the "A" insertion mutation. A new ssODN was also used with this strategy to insert the loxP sequence at the correct location, and to correct the "A" insertion mutation. To generate zygotes to be used for the creation of the floxed *COL4A1* containing mouse strain, sperm from a 3'loxP positive N1F1 male, oocytes from 3'loxP positive N1F1 and N2F1 females, and oocytes from WT C57BL/6J mice were used for IVF. Electroporation or microinjection were then used to insert the CRISPR-Cas9 components into the mouse zygotes at the one-cell stage, and the embryos were then implanted into pseudo pregnant female mice.

Using strategy one, 10 mice were produced, all which were screened by the three screening PCRs. Agarose gel results of the 5'loxP PCR showed four potential positives, with the most promising mouse appearing to have a homozygous insertion of the 5'loxP sequence (Figure 9A). Results of the agarose gel of the 3'loxP PCR showed five potential positives; however, the only mouse to appear to be positive for the 5'loxP and 3'loxP insertions was mouse 7469 (Figure 9B). To confirm that the 5' and 3' loxP insertions are on the same allele for mouse 7469, the loxP-loxP PCR was performed; however, a new forward primer was used in this PCR to account for

the new location of the 5'loxP insertion. The results of this PCR showed no amplification on the agarose gel; however, this PCR also lacks a positive control, and thus the results of this PCR were deemed inconclusive (Figure 9C). To gain confirmation about the possible loxP insertions, the 5'loxP region of the four mice appearing to be positive for the 5'loxP were sent for Sanger sequencing, and the 3'loxP region of mouse 7469 was sent for Sanger sequencing. Sanger sequencing results of the 5'loxP region showed that only mice 7464 and 7469 were true positives of the loxP insertion, with 7464 being heterozygous for the insertion and 7469 being homozygous for the loxP insertion (Figures 9D, 9E). However, sequencing results also revealed that the "A" insertion mutation has been corrected in mouse 7464, but not corrected in mouse 7469 (Figure 9F, 9G). Moreover, Sanger sequencing results of the 3'loxP region of mouse 7469 showed an interesting, yet confusing, result. Sequencing results showed that instead of having the 3'loxP insertion, the mouse contained a WT allele and a 2 base pair deletion allele (Figure 9H).

In total, 13 pups were born from strategy two, in which all pups were screened using the three screening PCRs. The 5'loxP PCR agarose gel results did not show any obvious possible positives like strategy one. Nonetheless, four mice were chosen for further screening via Sanger sequencing as they were determined to be the mice most likely to have the 5'loxP insertion (Figure 10A). The results of the 3'loxP PCR agarose gel showed six positive mice, with all four of the possible 5'loxP positive mice also being positive for the 3'loxP (Figure 10B). However, the loxP-loxP PCR agarose gel did not show any positive results (Figure 10C). The accuracy of the results of this PCR was then confirmed when Sanger sequencing of the 5'loxP region of the four possible positive mice revealed that none of the mice were true positives for the 5'loxP insertion (data not shown).

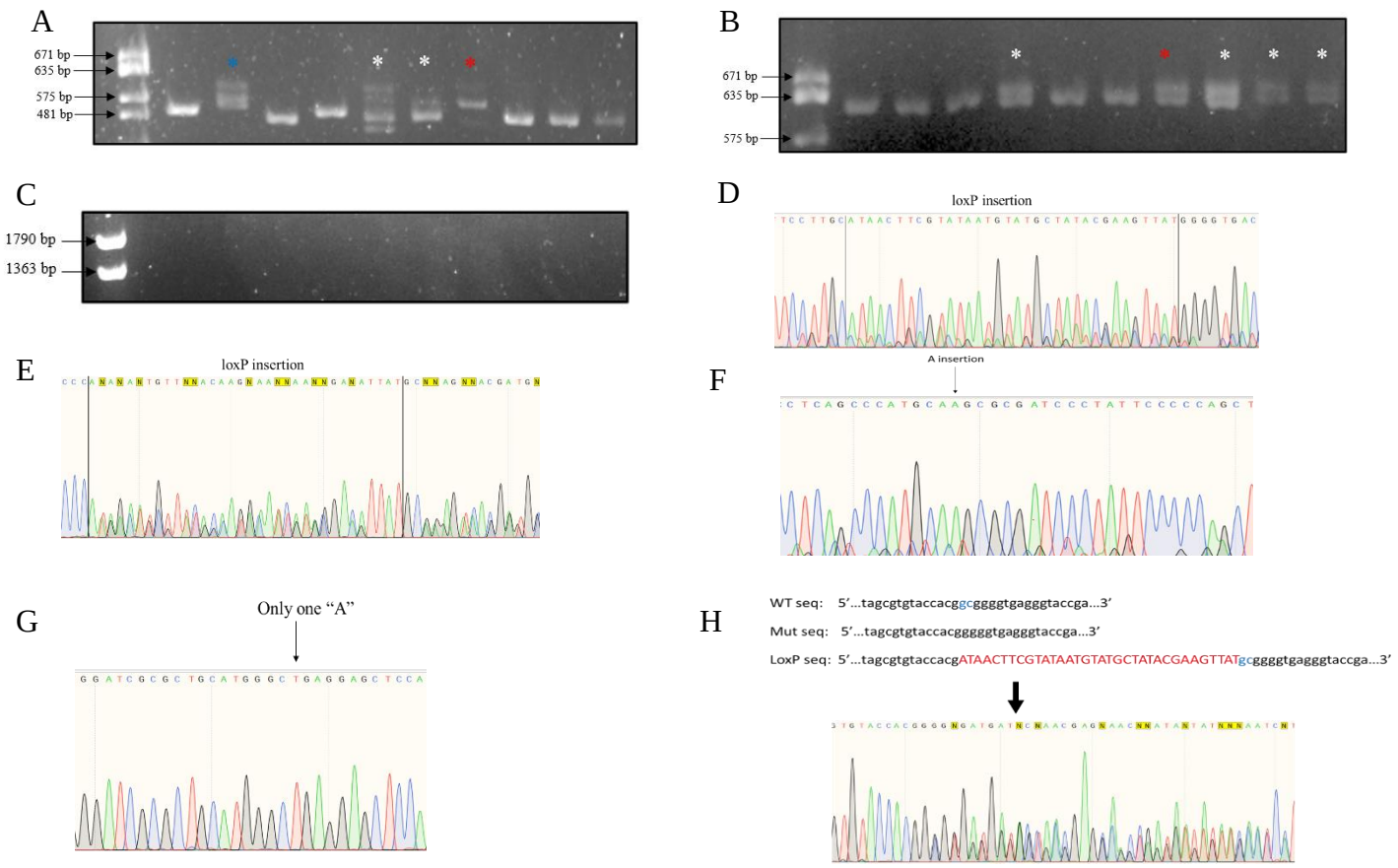

Figure 9. Screening of floxed *COL4A1* strategy one mice. (A) 5'loxP PCR results resolved by electrophoresis on 3% agarose gel. Positive mouse 7469 is indicated by a red asterisk, positive mouse 7464 is indicated by a blue asterisk, and remaining "positives" are indicated by a white asterisk. **(B)** 3'loxP PCR results resolved by electrophoresis on 3% agarose gel. Positive mouse 7469 is indicated by a red asterisk and remaining positives are indicated by a white asterisk. **(C)** loxP-loxP PCR results resolved by electrophoresis on 2% agarose gel. **(D)** Chromatogram from the Sanger sequencing results, showing the 5'loxP insertion for mouse 7469. **(E)** Chromatogram from the Sanger sequencing results, showing the heterozygous 5'loxP insertion for mouse 7464. **(F)** Chromatogram from the Sanger sequencing results, showing that the A insertion mutation is still present in mouse 7469. **(G)** Chromatogram from the Sanger sequencing results, showing that the A insertion mutation is not present in mouse 7464. Reverse primer was used to sequence, and thus "A" is a "T" for this sequence. **(H)** Diagram illustrating the Sanger sequencing results for the 3'loxP region in mouse 7469. Blue sequence present in the WT sequence and loxP sequence indicates the two base pairs that were deleted in one allele, the mutant sequence. Below is the chromatogram from the Sanger sequencing results, showing the sequencing results for the 3'loxP region in mouse 7469. Results indicate mouse 7469 is positive for 5'loxP and possibly positive for 3'loxP, and thus is most likely a positive founder mouse, but still has "A" insertion. Results indicate mouse 7464 is only positive for 5'loxP but does not have "A" insertion mutation.

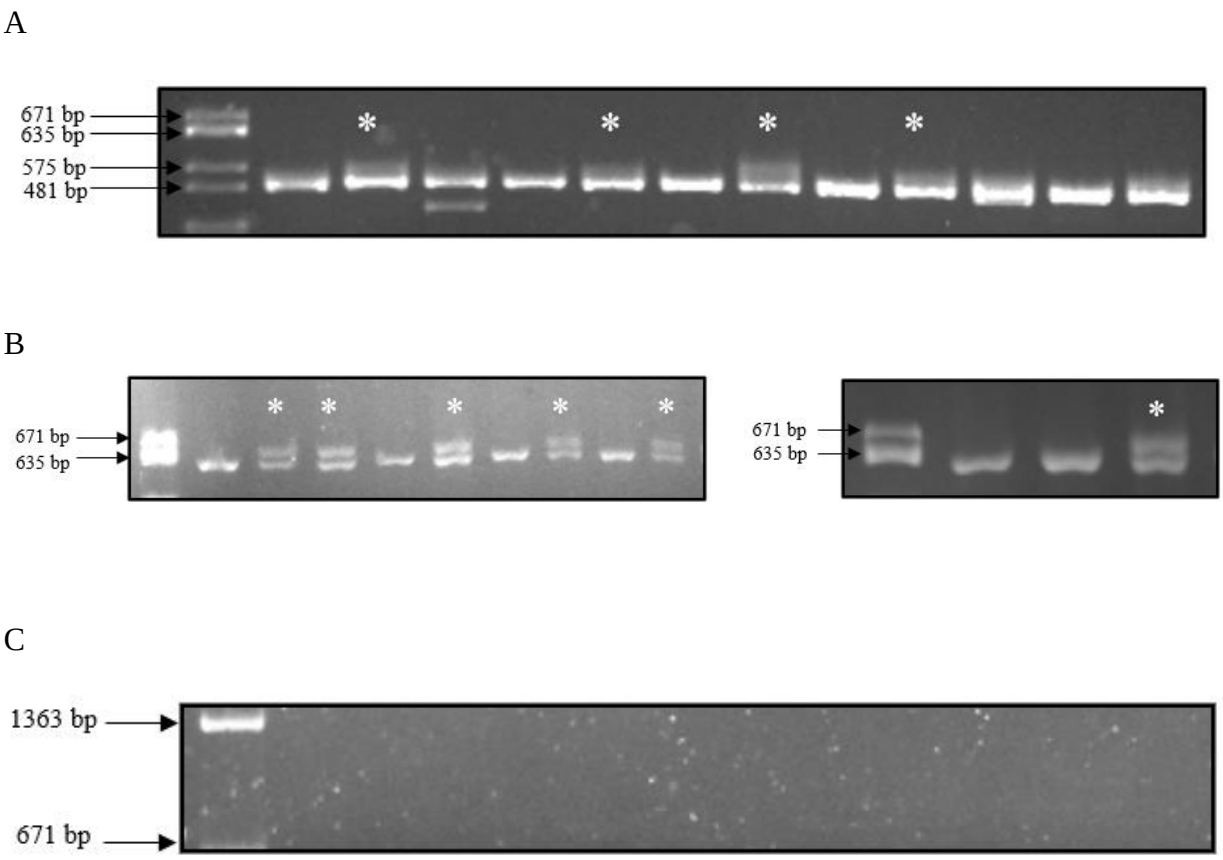

Figure 10. Screening of floxed *COL4A1* strategy two mice. (A) 5'loxP PCR results resolved by electrophoresis on 3% agarose gel. Samples believed to be possibly positive are indicated by white asterisk. **(B)** 3'loxP PCR results resolved by electrophoresis on 3% agarose gel. Positive samples are indicated by white asterisk. **(C)** loxP-loxP PCR results resolved by electrophoresis on 2% agarose gel.

4. DISCUSSION

4.1 Exploring the potential of Arresten to be used as a biomarker for cancer

Previous studies have investigated the connection between the tumour suppressor p53 and Arresten [50]. These previous studies have shown that p53 is able to induce expression of Arresten through several mechanisms. Furthermore, because p53 is mutated in approximately half of all cancers, the link between Arresten and p53 may prove to be interesting. If Arresten levels change in plasma levels of patients with certain types of cancer, then Arresten could potentially be used as a biomarker for detection of certain types of cancer. In order to investigate Arresten as a potential biomarker, *in vitro* experiments were first studied to show that secreted Arresten levels are a direct consequence of p53 action, and to show that Arresten levels can be quantified using ELISA. The results of these *in vitro* experiments help confirm that one-way p53 mediates its tumour suppressor function is by upregulating Arresten expression, and that Arresten levels are a direct consequence of p53 action. In particular, it is observed that greater p53 expression results in a stronger induction of Arresten expression, thus further confirming Arresten levels are a direct consequence of p53 action. Although this particular ELISA kit is not a clinically relevant tool for measurement of Arresten concentrations, the results nonetheless demonstrate the potential of secreted Arresten to be measured which is useful for research investigations into screening Arresten as a biomarker.

Once it was demonstrated that ELISA could be used as a tool to measure secreted Arresten concentrations, Arresten concentrations in human plasma samples from NSCLC patients could be measured to determine if Arresten has potential to be used as a biomarker for cancer detection. Results from the ELISA analysis showed that plasma samples from adenocarcinoma patients have statistically significant lower levels of Arresten compared to control patients with

no cancer at $p<0.1$. This indicates that further investigation should be conducted for using Arresten as a biomarker to detect adenocarcinoma via measurement of Arresten in plasma samples. Although, the results were only statistically significant at $p<0.1$, this could be due to several factors. Firstly, this study was only a preliminary study, and thus the sample size of both the adenocarcinoma samples and the control samples are not large enough to draw definitive conclusions or have greater statistical significance. Additionally, there may be other variables that could further stratify adenocarcinoma plasma samples to result in a more statistically significant decrease in Arresten concentration compared to control samples.

Furthermore, a major limitation of this study was a possible defect in the ELISA plates. There were very low readings across the plate even for the standards using this batch of ELISA plates which were a different batch compared to previous ELISA experiments performed. Literature searches indicate that minimal colour development may be due to errors in preparation of the plate, such as using the wrong concentration of capture antibody to coat the wells, the plate having gone bad, or the standard having gone bad [84]. Thus, any conclusions drawn from this study may not be accurate and this study may not be reproducible.

The next step of this study should be to repeat the experiment with the same samples using ELISA plates from a different lot number to confirm the reproducibility. If the results are repeated, a study with a larger sample size of adenocarcinoma samples and control samples should be conducted. Further investigation could then be performed to stratify adenocarcinoma plasma samples based on additional variables such as grade of cancer or stage of cancer, to determine if Arresten levels change with other variables. Additionally, the previous study in our lab suggested that over long periods of time, Arresten plasma levels do not change significantly in NSCLC patients over time. Thus, it could be interesting to examine Arresten concentrations in

adenocarcinoma plasma samples compared to control samples over an extended period of time determine if this phenomenon is specifically observed in adenocarcinoma samples. If it is, this could suggest that low Arresten levels predispose people to developing adenocarcinoma, and thus Arresten levels could be a way to screen for high probability of developing adenocarcinoma.

4.2 Arresten may be processed from COL4A1 via a secreted protease that is inhibited by GM6001 by not marimastat

The processing of Arresten from the ECM is not well understood. Previous studies have suggested that MMPs, namely MMP-2, may play a role in the processing as p53 induces the expression of MMP-2 and may induce the expression of other MMPs [52]. Furthermore, studies have shown that other collagen-derived matrikines are processed by MMPs [53]. Therefore, in order to investigate the processing of Arresten, activity of different MMPs was inhibited to determine if any MMPs play a role in the processing. Using the broad-spectrum MMP inhibitors, GM6001 and marimastat, it was shown that MMPs inhibited by GM6001 but not marimastat may play a role in the processing of Arresten. However, it should be noted that no assay was performed to ensure that the MMP inhibitors successfully inhibited their targets. As a result, it is possible that the concentration of marimastat used was not the optimal to inhibit the MMPs, and the target MMPs were, in fact, not inhibited. If this experiment were to be repeated, it would be beneficial to perform an additional assay, such as a gelatin zymography assay to detect MMPs in the conditioned media, to confirm that both MMP inhibitors are working.

Previous studies have reported that the proteases MMP-3, MMP-9, MMP-12, MMP- 26, and ADAM-10 are inhibited by GM6001 but not Marimastat and are expressed in H1299 cells [85], [86]. Therefore, to try and elucidate the protease that is responsible for Arresten processing, the

five proteases should be individually knocked out in H1299 cells. We had previously attempted to create knockout cell lines for all five proteases in H1299 cells via lentiviral infection of CRISPR-Cas9 components with a population of H1299 cells. However, this method was not successful in creating the different knockout cell lines as it instead produced a heterogeneous cell population, in which some cells may have had the target escape modification, or some cells may have had the target protein modified but still able to function [87]. Instead, CRISPR-edited clones derived should be used to generate the knockout cell lines [88]. Once the knockout cell lines have been generated, they could be infected with Ad-p53 to induce p53, and subsequently Arresten, expression. The conditioned media of these cells could then be assessed for Arresten expression via western blot or ELISA to determine how knocking out a specific secreted protease affects Arresten expression, and thus to determine if the specific protease is responsible for Arresten processing.

4.3 Creating a floxed *COL4A1* mouse strain

Using CRISPR-Cas9 has shown great promise in easing the process of creating floxed mouse strains compared to historically traditional methods. However, from observing the attempts to create a floxed *COL4A1* mouse strain, it is clear that further optimization of this process and the process for screening founders is required in order to streamline the process of creating a floxed mouse strain.

In order to select sgRNAs to use to insert the upstream and downstream loxP sequences, sgRNA testing was performed. The results of gRNA testing suggest that this process is essential and highly beneficial in terms of streamlining the overall process of creating a floxed mouse strain. There are currently many sgRNA design tools which are free of use on the internet, and researchers may be tempted to simply select the sgRNAs that receive the best score according to

these online programs. However, guide efficiency testing of the upstream sgRNAs showed that sgRNAs predicted to have the highest scores may not truly be the most efficient sgRNA to use. It was observed that while "gRNA up 1" was predicted to be the most efficient and "gRNA up 3" was predicted to be the least efficient, "gRNA up 3" was actually the most efficient, whereas "gRNA up 1" was the least efficient. Therefore, it is essential to perform sgRNA efficiency testing to ensure that appropriate sgRNAs are selected, and that researchers do not wait until screening founders to realize that the sgRNAs selected are not efficient.

Screening of potential founders using PCR to amplify different regions and then using Sanger sequencing to sequence these regions proved to be somewhat of a laborious and ineffective process that did not yield any concrete results, due to Sanger sequencing producing chromatograms that contained many sequences and were impossible to interpret to determine if the loxP sequence had been inserted. This is due to the fact that many of the mice are mosaic because CRISPR-Cas9 components were injected in two-cell stage embryos. At the two-cell stage, the first genomic DNA replication has most likely already occurred, and therefore genetic editing occurs after the first replication, resulting in mosaicism [89]. One way to combat the issue of mosaicism in the mice, and thus ensure that screening for potential founders is more effective, is to inject the CRISPR-Cas9 components in one cell stage embryos because then there is less chance of integration occurring after the first genomic DNA replication. However, previous studies have shown that insertion of the CRISPR-Cas9 components at the two-cell stage may be more efficient than injection at the one-cell stage [90]. In order for the loxP sequence to be inserted at the target location, homologous recombination using the inserted repair template must occur at the cut site of the DNA. Homologous recombination is the most active at the late S-G2 phases of the cell cycle, and at the two-cell stage for mouse embryos, the G2 phase is long,

about 10-12 hours [90]. Furthermore, at the two-cell stage at the G2 phase, the chromatin is in an open state, thus increasing the accessibility of the chromatin to Cas9 and the repair template [90]. Therefore, a longer G2 phase with increased accessibility of the chromatin is likely to result in more successful homologous recombination, and as a result, it may be more beneficial to inject the CRISPR-Cas9 components in two-cells stage embryos, despite the mosaicism that arises in the mice.

If it is more beneficial to inject the CRISPR-Cas9 components in two-cell stage embryos, then potential founder mice will most likely be mosaic, and Sanger sequencing would still not be ideal for screening for founder mice but sufficient to screen F1 and subsequent generations. This is because Sanger sequencing is unable to detect low-level mosaic alleles present in founder mice but not present in F1 and subsequent generations [91]. Alternatively, next-generation sequencing, such as Illumina's MiSeq, may prove to be a more effective sequencing method for screening founders. MiSeq is a high-throughput deep sequencing technology, which sequences by synthesis, and is ideal for targeted gene sequencing [92]. This technology may be ideal for identifying the different sequences present in the mosaic regions of the DNA as this technology is able to sequence from single DNA fragments and therefore is able to sequence low-frequency variants [93]. As a result, this sequencing technology could allow for easier identification of whether the loxP sequence was inserted or not. Additionally, screening results of founder mice revealed that although one mouse, mouse 5478, was a likely founder candidate, this mouse was not able to transmit the 5'loxP mutation to the F1 generation. Thus, the 5'loxP insertion was not a germline mutation, but rather a somatic mutation. MiSeq could also provide a way to predict whether the loxP insertions, or other mutations, are germline or somatic. Although MiSeq cannot actually determine whether a mutation is somatic or germline, it can give information about the

frequency of the loxP insertion [94]. Therefore, if MiSeq results reveal that the loxP insertion occurred at a low frequency, it could be predicted that the mutation is not likely going to be in the germline, and thus will not be passed down to the F1 generation. This would allow for more efficient screening of founder candidates to determine which should be mated with WT mice to produce F1 litters, as this information would help choose which mice have the loxP insertion at a higher frequency and are thus more likely to pass the mutation to their F1 offspring.

There are also other possible ways to improve the efficiency of CRISPR-Cas9 and homologous recombination to result in increased insertion of the loxP sequences. A study was conducted by Chen et al. (2011), where genome editing using ssODNs and zinc-finger nucleases was investigated. This study showed that using ssODNS to mediate knock in of a desired sequence is more efficient than using a double stranded donor template [76]. Therefore, this study may be applicable to generating conditional knockout mice using CRISPR, and it would be possible to increase efficiency of loxP insertions by using a ssODN instead of a double stranded repair template. Additionally, the efficiency of using double-stranded breaks and HDR to introduce DNA fragments into desired locations is greatly limited by NHEJ which competes with HDR. Therefore, to increase the frequency of HDR and decrease the frequency of NHEJ, a study was performed to investigate the effects on HDR-mediated genome editing by inhibiting the NEHJ pathway [95]. In particular, the effect of transiently inhibiting DNA ligase IV, an enzyme in the NHEJ pathway, using the inhibitor Scr7 in mice was tested, where Scr7 was co-injected with the CRISPR-Cas9 components into mouse zygotes. The study showed that co-injection Scr7 with the CRISPR-Cas9 components significantly increases HDR-mediated insertional mutagenesis compared to when only the CRISPR-Cas9 components are injected, with no effects on zygote viability or number of live pups born. Therefore, it would be beneficial to also co-

inject Scr7 with the CRISPR-Cas9 components to increase HDR-mediated insertional mutagenesis when generating conditional knockout mice. Lastly, some studies have shown that electroporation may be more efficient for producing genetic modifications in mice compared to microinjection. In particular, Kaneko (2017) showed that using the Technique for Animal Knockout system by Electroporation (TAKE) to generate genome-edited rats using CRISPR-Cas9 results in a greater percentage of genetically modified offspring compared to when microinjection is used [96]. Specifically, when using TAKE, 100% of the offspring born contained the desired genetic modification, whereas only 77% of the offspring contained the desired genetic modification when using microinjection.

Sanger sequencing results of the region where the 5'loxP should be inserted in the 3'loxP positive F1 generation showed that the mice were heterozygous for an "A" insertion mutation within the region of where the upstream sgRNA was supposed to anneal. Additionally, Sanger sequencing results of this region in purchased WT C57BL/6J mice showed no "A" insertion mutation in these mice, whereas Sanger sequencing results of other potential founder mice showed them to be homozygous for this mutation. As a result, it was concluded that the F1 mice inherited this "A" insertion mutation from the founder father, mouse 5478. This led to the hypothesis that the reason why so few potential founder mice had the 5'loxP sequence inserted was because this A insertion mutation prevented the upstream sgRNA from binding stably to its target location, as this "A" insertion mutation was within the annealing region of the sgRNA. This then reduced cleavage of the DNA by Cas9. Furthermore, the "A" insertion mutation was within the PAM-proximal region of the sgRNA. Previous studies have shown that mismatches between the sgRNA and DNA in the PAM-proximal region of the sgRNA is less tolerated by Cas9 than mismatches in the PAM-distal region of the sgRNA [97]. It has been suggested that

this is because the 8-12 PAM-proximal bases of the sgRNA make contact with the recognition lobe of Cas9 [71]. Therefore, these studies help reinforce our hypothesis that mismatch between the sgRNA and target DNA due to the "A" insertion mutation may have reduced cleavage of the DNA and thus resulted in low frequency insertion of the upstream loxP sequence in the first injection round of mice. Based on these results, it may be beneficial in the future to use Sanger sequencing to screen the region of interest in the mice whose sperm and oocytes will be used for IVF to generate the genetically modified mouse strain to ensure there are no mutations in the regions where the sgRNAs will anneal.

Screening of the floxed *COL4A1* founder mice produced using strategy one to insert the 5'loxP sequence showed that the mouse 7469 is the most likely candidate to have the 5'loxP and 3'loxP. Both the 5'loxP and 3'loxP agarose gels were positive for mouse 7469, however, Sanger sequencing only confirmed the 5'loxP insertion. Furthermore, the loxP-loxP PCR was negative for mouse 7469. However, the loxP-loxP PCR was deemed inconclusive because of the lack of positive control for this PCR. The lack of positive control for this PCR prevented optimization of the PCR reaction components and PCR running and cycling parameters. As a result, it cannot be determined whether the PCR reaction did not work due to the PCR reaction components and parameters not being optimal to produce amplification of the region or did not work due to mouse 7469 not being a true positive for both the 5'loxP and 3'loxP insertions on the same allele. Additionally, Sanger sequencing results of the 3'loxP region of mouse 7469 showed that instead of the downstream loxP sequence being inserted, mouse 7469 contained a WT allele and an allele containing a 2 base pair deletion at the site where the loxP sequence should have been inserted. This sequencing result does not correlate with what is observed on agarose gel, as on gel a band corresponding to the size of the WT band is observed, and a band larger than the WT

band is observed. This would indicate that the mouse is heterozygous for some type of insertion of bases in this region, not a deletion of bases like the sequencing results suggest. As a result, it is possible that Sanger sequencing results do not accurately reflect the sequence of the 3'loxP region of mouse 7469.

Furthermore, screening results of the floxed *COL4A1* founder mice produced using strategy one and two to insert the 5'loxP sequence suggested that strategy one is a more efficient strategy than strategy two. As mentioned earlier, the sgRNA used in strategy one had been previously tested whereas the sgRNA used in strategy two had not been tested but was a slightly modified version of the sgRNA that had been for the first round of injections with the double-stranded repair template. This result contradicts what previous studies would predict to be the most efficient sgRNA. While investigating sequence features that improve sgRNA design, Xu et al. (2014) suggest that different sequence features contribute to sgRNA efficiency [98]. First, they suggest that adenines are preferred at positions -5 to -12 of the sgRNA. The sgRNA used in strategy two has two adenines within this range, whereas the sgRNA used in strategy one does not contain any adenines. Secondly, they suggest that guanines are preferred at positions -14 to -17 of the sgRNA. Again, the sgRNA used in strategy two has one guanine within this range, whereas the sgRNA used in strategy one does not contain any guanines in this range. Moreover, in a study conducted by Doench et al., (2014) to also investigate features of highly active sgRNAs, they suggest that there is a strong preference for cytosine at position 16 of the sgRNA, and that sgRNAs composed of moderate GC content should make the sgRNA highly efficient [99]. Although both features are present in both sgRNAs, the sgRNA used in strategy one was more efficient than the sgRNA used in strategy two. Therefore, these results further stress the importance of sgRNA testing when using CRISPR-Cas9 to create genetically modified mice, as

even consideration of various sequence features of sgRNAs cannot accurately predict the efficiency of sgRNAs.

In the short-term future, first, it would be ideal to confirm whether the 3'loxP sequence is truly present in mouse 7469 or not. This could be achieved by using PCR to amplify the region again and then using Sanger sequencing to sequence the region. If it is confirmed that the 3'loxP sequence is not present in mouse 7469, then mouse 7464 could be used to insert the 3'loxP sequence. Mouse 7464 could be mated with a WT male to produce F1 litters, and any 5'loxP positive females and males from the F1 litters could then be used for IVF to produce zygotes that contain the 5'loxP insertion. The zygotes could then be microinjected or electroporated with CRISPR-Cas9 components and an ssODN to insert the 3'loxP, in a similar manner in which the floxed *COL4A1* mice using strategy one and two to insert the 5'loxP sequence were created. Moreover, even if it is confirmed that mouse 7469 does contain the 3'loxP sequence, it may still be worthwhile to use oocytes and sperm from F1 litters from mouse 7464 to insert the 3'loxP sequence. Mouse 7464 does not contain the "A" insertion mutation, which makes it a more ideal founder. Additionally, a novel mouse line should have multiple founders which could be achieved by using F1 progeny to insert the 3'loxP sequence. Eventually, when the floxed *COL4A1* mouse strain is established, the mice will need to be crossed with a whole body Cre-ER^{T2} mouse line to produce a mouse strain in which *COL4A1* can be conditionally deleted throughout the whole adult mouse via injection of tamoxifen to induce Cre expression. When *COL4A1* is conditionally deleted throughout a whole adult mouse, there are several experiments that can be done to investigate Arresten as a biomarker for cancer detection, and to characterize the effect of Arresten on the tumour microenvironment. Firstly, ELISA could be used to characterize basal Arresten levels in the plasma of these mice compared to WT mice. Although,

the goal is to create a whole-body knockout of *COL4A1*, it is possible, that complete removal of *COL4A1* and *COL4A2* in adult mice will result in lethality. However, because a tamoxifen inducible Cre system is being used, the levels of tamoxifen given to the mouse can be varied in order to cause less efficient excision of the critical exon. Thus, some COL4A1 may still be produced, and by extension there may still be some circulating levels of arresten in the plasma. However, arresten levels should be reduced, and this reduction could be characterized using ELISA. Lastly, certain syngeneic tumour lines can be implanted in the mice, to determine how reducing arresten levels affects tumour growth, angiogenesis, and metastasis in non-immunocompromised mice. This would also allow for continued characterization of Arresten as a biomarker for cancer detection.

4.4 Summary

Overall, this thesis helps to further characterize Arresten and aid in the process of creating a floxed *COL4A1* mouse strain. Arresten protein levels are a direct consequence of p53 action, and because p53 is mutated in about half of all cancers, this may have an effect of Arresten protein levels and thus Arresten could be used as a biomarker for cancer detection. Investigations into Arresten concentrations in plasma samples of the different histological subtypes of NSCLC suggested that Arresten concentrations may be lower in the plasma of patients with adenocarcinoma compared to patients with no cancer. However, further investigation should be done to ensure that the results are reproducible due to issues with observed with the ELISA kits, and to ensure that the results are reproducible with a larger sample size. Further investigation should also be done to determine if Arresten levels change based on other variables, or to determine how or if Arresten levels change over a long period of time. This thesis also demonstrates that Arresten may be processed from COL4A1 by secreted proteases that are

inhibited by GM6001 but not marimastat. Further experiments should be done to first validate that the concentrations of GM6001 and marimastat used do in fact inhibit the targeted secreted proteases. If this is validated, then knockout cell lines of the target proteases inhibited by GM6001 but not marimastat should be created to ascertain exactly which proteases are responsible for processing Arresten. Lastly, although no confirmation was obtained regarding whether a floxed *COL4A1* founder mouse was created, it seems likely that a floxed *COL4A1* founder mouse was created or would soon be created. Additionally, this thesis helped to demonstrate flaws in the current CRISPR-Cas9 process used to create floxed mice strains, and potential ways in which this process could be optimized.

5. REFERENCES

1. Wang, M., et al., *Role of tumor microenvironment in tumorigenesis.* Journal of Cancer, 2017. **8**(5): p. 761-773.
2. Lu, P., V.M. Weaver, and Z. Werb, *The extracellular matrix: a dynamic niche in cancer progression.* The Journal of cell biology, 2012. **196**(4): p. 395-406.
3. Kalluri, R., *Basement membranes: structure, assembly and role in tumour angiogenesis.* Nature Reviews Cancer, 2003. **3**(6): p. 422-433.
4. Tran, K.T., P. Lamb, and J.-S. Deng, *Matrikines and matricryptins: Implications for cutaneous cancers and skin repair.* Journal of Dermatological Science, 2005. **40**(1): p. 11-20.
5. Maquart, F.-X., et al., *An introduction to matrikines: extracellular matrix-derived peptides which regulate cell activity: Implication in tumor invasion.* Critical Reviews in Oncology/Hematology, 2004. **49**(3): p. 199-202.
6. Boraschi-Diaz, I., et al., *Collagen Type I as a Ligand for Receptor-Mediated Signaling.* Frontiers in Physics, 2017. **5**(12).
7. Ricard-Blum, S., *The collagen family.* Cold Spring Harbor perspectives in biology, 2011. **3**(1): p. a004978-a004978.
8. Manon-Jensen, T., N.G. Kjeld, and M.A. Karsdal, *Collagen-mediated hemostasis.* Journal of Thrombosis and Haemostasis, 2016. **14**(3): p. 438-448.
9. Leitinger, B. and E. Hohenester, *Mammalian collagen receptors.* Matrix Biology, 2007. **26**(3): p. 146-155.
10. Exposito, J.-Y., et al., *The Fibrillar Collagen Family.* International Journal of Molecular Sciences, 2010. **11**(2).
11. Kisling, A., R.M. Lust, and L.C. Katwa, *What is the role of peptide fragments of collagen I and IV in health and disease?* Life Sciences, 2019. **228**: p. 30-34.
12. Persikov, A.V., et al., *Electrostatic Interactions Involving Lysine Make Major Contributions to Collagen Triple-Helix Stability.* Biochemistry, 2005. **44**(5): p. 1414-1422.
13. Khoshnoodi, J., et al., *Molecular Recognition in the Assembly of Collagens: Terminal Noncollagenous Domains Are Key Recognition Modules in the Formation of Triple Helical Protomers *.* Journal of Biological Chemistry, 2006. **281**(50): p. 38117-38121.
14. Hynes, R.O., *The Extracellular Matrix: Not Just Pretty Fibrils.* Science, 2009. **326**(5957): p. 1216.
15. Fox, M.A., *Novel roles for collagens in wiring the vertebrate nervous system.* Current Opinion in Cell Biology, 2008. **20**(5): p. 508-513.
16. Cheng, J.S., et al., *Collagen VI protects neurons against Aβ toxicity.* Nature Neuroscience, 2009. **12**(2): p. 119-121.
17. Ricard-Blum, S. and S.D. Vallet, *Fragments generated upon extracellular matrix remodeling: Biological regulators and potential drugs.* Matrix Biology, 2019. **75-76**: p. 170-189.
18. Nyberg, P., L. Xie, and R. Kalluri, *Endogenous Inhibitors of Angiogenesis.* Cancer Research, 2005. **65**(10): p. 3967.
19. O'Reilly, M.S., et al., *Endostatin: An Endogenous Inhibitor of Angiogenesis and Tumor Growth.* Cell, 1997. **88**(2): p. 277-285.

20. Walia, A., et al., *Endostatin's emerging roles in angiogenesis, lymphangiogenesis, disease, and clinical applications.* Biochimica et biophysica acta, 2015. **1850**(12): p. 2422-2438.
21. Xu, X., et al., *Endostar, a Modified Recombinant Human Endostatin, Suppresses Angiogenesis through Inhibition of Wnt/β-Catenin Signaling Pathway.* PLOS ONE, 2014. **9**(9): p. e107463.
22. Monboisse, J.C., et al., *Matrikines from basement membrane collagens: A new anti-cancer strategy.* Biochimica et Biophysica Acta (BBA) - General Subjects, 2014. **1840**(8): p. 2589-2598.
23. Ruge, T., et al., *Endostatin: a promising biomarker in the cardiovascular continuum?* Biomarkers in Medicine, 2017. **11**(10): p. 905-916.
24. Zhang, C., et al., *Endostatin as a novel prognostic biomarker in acute ischemic stroke.* Atherosclerosis, 2019. **293**.
25. Holm Nielsen, S., et al., *Tumstatin, a Matrikine Derived from Collagen Type IVα3, is Elevated in Serum from Patients with Non-Small Cell Lung Cancer.* Translational oncology, 2018. **11**: p. 528-534.
26. Ohlund, D., et al., *Type IV collagen is a tumour stroma-derived biomarker for pancreas cancer.* British journal of cancer, 2009. **101**(1): p. 91-97.
27. Momota, R., et al., *Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and alpha4(IV) collagen chains are arranged head-to-head on chromosome 2q36.* FEBS Lett, 1998. **424**(1-2): p. 11-6.
28. Zhou, J., et al., *Deletion of the paired alpha 5(IV) and alpha 6(IV) collagen genes in inherited smooth muscle tumors.* Science, 1993. **261**(5125): p. 1167-9.
29. Kamphaus, G.D., et al., *Canstatin, a Novel Matrix-derived Inhibitor of Angiogenesis and Tumor Growth*.* Journal of Biological Chemistry, 2000. **275**(2): p. 1209-1215.
30. Kumar, R., et al., *Spatial and temporal expression of angiogenic molecules during tumor growth and progression.* Oncol Res, 1998. **10**(6): p. 301-11.
31. Folkman, J., *Angiogenesis in cancer, vascular, rheumatoid and other disease.* Nature Medicine, 1995. **1**(1): p. 27-30.
32. Holash, J., S.J. Wiegand, and G.D. Yancopoulos, *New model of tumor angiogenesis: dynamic balance between vessel regression and growth mediated by angiopoietins and VEGF.* Oncogene, 1999. **18**(38): p. 5356-5362.
33. Rak, J., et al., *Oncogenes and Angiogenesis: Signaling Three-Dimensional Tumor Growth.* Journal of Investigative Dermatology Symposium Proceedings, 2000. **5**(1): p. 24-33.
34. Carmeliet, P., et al., *Role of HIF-1α in hypoxia-mediated apoptosis, cell proliferation and tumour angiogenesis.* Nature, 1998. **394**(6692): p. 485-490.
35. Fukumura, D., et al., *Tumor Induction of VEGF Promoter Activity in Stromal Cells.* Cell, 1998. **94**(6): p. 715-725.
36. Lyden, D., et al., *Impaired recruitment of bone-marrow-derived endothelial and hematopoietic precursor cells blocks tumor angiogenesis and growth.* Nat Med, 2001. **7**(11): p. 1194-201.
37. Colorado, P.C., et al., *Anti-angiogenic cues from vascular basement membrane collagen.* Cancer Res, 2000. **60**(9): p. 2520-6.
38. Aikio, M., et al., *Arresten, a collagen-derived angiogenesis inhibitor, suppresses invasion of squamous cell carcinoma.* PloS one, 2012. **7**(12): p. e51044-e51044.

39. Sudhakar, A. and C.S. Boosani, *Signaling mechanisms of endogenous angiogenesis inhibitors derived from type IV collagen.* Gene regulation and systems biology, 2007. **1**: p. 217-226.
40. Boosani, C.S., et al., *FAK and p38-MAP Kinase-Dependent Activation of Apoptosis and Caspase-3 in Retinal Endothelial Cells by α1(IV)NC1.* Investigative Ophthalmology & Visual Science, 2009. **50**(10): p. 4567-4575.
41. Levine, A.J., J. Momand, and C.A. Finlay, *The p53 tumour suppressor gene.* Nature, 1991. **351**(6326): p. 453-456.
42. Zhang, X.-d., Z.-h. Qin, and J. Wang, *The role of p53 in cell metabolism.* Acta Pharmacologica Sinica, 2010. **31**(9): p. 1208-1212.
43. Teodoro, J.G., S.K. Evans, and M.R. Green, *Inhibition of tumor angiogenesis by p53: a new role for the guardian of the genome.* J Mol Med (Berl), 2007. **85**(11): p. 1175-86.
44. Ho, J. and S. Benchimol, *Transcriptional repression mediated by the p53 tumour suppressor.* Cell Death & Differentiation, 2003. **10**(4): p. 404-408.
45. Ravi, R., et al., *Regulation of tumor angiogenesis by p53-induced degradation of hypoxia-inducible factor 1alpha.* Genes & development, 2000. **14**(1): p. 34-44.
46. Krock, B.L., N. Skuli, and M.C. Simon, *Hypoxia-induced angiogenesis: good and evil.* Genes & cancer, 2011. **2**(12): p. 1117-1133.
47. Kubbutat, M.H., S.N. Jones, and K.H. Vousden, *Regulation of p53 stability by Mdm2.* Nature, 1997. **387**(6630): p. 299-303.
48. Chen, J., *The Cell-Cycle Arrest and Apoptotic Functions of p53 in Tumor Initiation and Progression.* Cold Spring Harbor perspectives in medicine, 2016. **6**(3): p. a026104-a026104.
49. Levine, A.J. and M. Oren, *The first 30 years of p53: growing ever more complex.* Nature Reviews Cancer, 2009. **9**(10): p. 749-758.
50. Assadian, S., et al., *p53 Inhibits Angiogenesis by Inducing the Production of Arresten.* Cancer research, 2012. **72**: p. 1270-9.
51. Teodoro, J.G., et al., *p53-Mediated Inhibition of Angiogenesis Through Up-Regulation of a Collagen Prolyl Hydroxylase.* Science, 2006. **313**(5789): p. 968.
52. Bian, J. and Y. Sun, *Transcriptional activation by p53 of the human type IV collagenase (gelatinase A or matrix metalloproteinase 2) promoter.* Molecular and cellular biology, 1997. **17**(11): p. 6330-6338.
53. Hamano, Y., et al., *Physiological levels of tumstatin, a fragment of collagen IV alpha3 chain, are generated by MMP-9 proteolysis and suppress angiogenesis via alphaV beta3 integrin.* Cancer cell, 2003. **3**(6): p. 589-601.
54. Poschl, E., et al., *Collagen IV is essential for basement membrane stability but dispensable for initiation of its assembly during early development.* Development, 2004. **131**(7): p. 1619-28.
55. Hall, B., A. Limaye, and A.B. Kulkarni, *Overview: generation of gene knockout mice.* Current protocols in cell biology, 2009. **Chapter 19**: p. Unit-19.12.17.
56. Rajewsky, K., et al., *Conditional gene targeting.* The Journal of clinical investigation, 1996. **98**(3): p. 600-603.
57. Sauer, B. and N. Henderson, *Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1.* Proceedings of the National Academy of Sciences of the United States of America, 1988. **85**(14): p. 5166-5170.

58. Sternberg, N. and D. Hamilton, *Bacteriophage P1 site-specific recombination: I. Recombination between loxP sites.* Journal of Molecular Biology, 1981. **150**(4): p. 467-486.
59. Gu, H., et al., *Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting.* Science, 1994. **265**(5168): p. 103.
60. Abremski, K., R. Hoess, and N. Sternberg, *Studies on the properties of P1 site-specific recombination: Evidence for topologically unlinked products following recombination.* Cell, 1983. **32**(4): p. 1301-1311.
61. Friedel, R.H., et al., *Generating Conditional Knockout Mice*, in *Transgenic Mouse Methods and Protocols*, M.H. Hofker and J. van Deursen, Editors. 2011, Humana Press: Totowa, NJ. p. 205-231.
62. Feil, R., et al., *Regulation of Cre Recombinase Activity by Mutated Estrogen Receptor Ligand-Binding Domains.* Biochemical and Biophysical Research Communications, 1997. **237**(3): p. 752-757.
63. Seibler, J., et al., *Rapid generation of inducible mouse mutants.* Nucleic acids research, 2003. **31**(4): p. e12-e12.
64. Schwenk, F., et al., *Temporally and spatially regulated somatic mutagenesis in mice.* Nucleic Acids Research, 1998. **26**(6): p. 1427-1432.
65. Jahn, H.M., et al., *Refined protocols of tamoxifen injection for inducible DNA recombination in mouse astroglia.* Scientific Reports, 2018. **8**(1): p. 5913.
66. Ishino, Y., et al., *Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product.* Journal of bacteriology, 1987. **169**(12): p. 5429-5433.
67. Adli, M., *The CRISPR tool kit for genome editing and beyond.* Nature Communications, 2018. **9**(1): p. 1911.
68. Barrangou, R., et al., *CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes.* Science, 2007. **315**(5819): p. 1709.
69. Bhaya, D., M. Davison, and R. Barrangou, *CRISPR-Cas Systems in Bacteria and Archaea: Versatile Small RNAs for Adaptive Defense and Regulation.* Annual Review of Genetics, 2011. **45**(1): p. 273-297.
70. Marraffini, L.A., *CRISPR-Cas immunity in prokaryotes.* Nature, 2015. **526**(7571): p. 55-61.
71. Nishimasu, H., et al., *Crystal structure of Cas9 in complex with guide RNA and target DNA.* Cell, 2014. **156**(5): p. 935-49.
72. Anders, C., et al., *Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease.* Nature, 2014. **513**(7519): p. 569-573.
73. Jinek, M., et al., *Structures of Cas9 endonucleases reveal RNA-mediated conformational activation.* Science (New York, N.Y.), 2014. **343**(6176): p. 1247997-1247997.
74. Jinek, M., et al., *A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity.* Science, 2012. **337**(6096): p. 816.
75. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nature Protocols, 2013. **8**(11): p. 2281-2308.
76. Chen, F., et al., *High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases.* Nature Methods, 2011. **8**(9): p. 753-755.
77. Miura, H., et al., *Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors.* Nature Protocols, 2018. **13**(1): p. 195-215.

78. Winer, A., S. Adams, and P. Mignatti, *Matrix Metalloproteinase Inhibitors in Cancer Therapy: Turning Past Failures Into Future Successes.* Molecular cancer therapeutics, 2018. **17**(6): p. 1147-1155.
79. Sado, Y., et al., *Establishment by the rat lymph node method of epitope-defined monoclonal antibodies recognizing the six different alpha chains of human type IV collagen.* Histochem Cell Biol, 1995. **104**(4): p. 267-75.
80. Concordet, J.P. and M. Haeussler, *CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens.* Nucleic Acids Res, 2018. **46**(W1): p. W242-W245.
81. Bunz, F., et al., *Requirement for p53 and p21 to sustain G2 arrest after DNA damage.* Science, 1998. **282**(5393): p. 1497-501.
82. Rudolf, E., K. Rudolf, and M. Cervinka, *Camptothecin induces p53-dependent and -independent apoptogenic signaling in melanoma cells.* Apoptosis, 2011. **16**(11): p. 1165.
83. Fields, G.B., *The Rebirth of Matrix Metalloproteinase Inhibitors: Moving Beyond the Dogma.* Cells, 2019. **8**(9): p. 984.
84. Aydin, S., *A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA.* Peptides, 2015. **72**: p. 4-15.
85. Metz, V.V., et al., *Induction of RAGE Shedding by Activation of G Protein-Coupled Receptors.* PLOS ONE, 2012. **7**(7): p. e41823.
86. Mondal, S., et al., *Matrix metalloproteinase-9 (MMP-9) and its inhibitors in cancer: A minireview.* European Journal of Medicinal Chemistry, 2020. **194**: p. 112260.
87. Sharpe, J.J. and T.A. Cooper, *Unexpected consequences: exon skipping caused by CRISPR-generated mutations.* Genome biology, 2017. **18**(1): p. 109-109.
88. Giuliano, C.J., et al., *Generating Single Cell-Derived Knockout Clones in Mammalian Cells with CRISPR/Cas9.* Current protocols in molecular biology, 2019. **128**(1): p. e100-e100.
89. Aoki, F., D.M. Worrad, and R.M. Schultz, *Regulation of Transcriptional Activity during the First and Second Cell Cycles in the Preimplantation Mouse Embryo.* Developmental Biology, 1997. **181**(2): p. 296-307.
90. Gu, B., E. Posfai, and J. Rossant, *Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos.* Nature Biotechnology, 2018. **36**(7): p. 632-637.
91. Jamuar, S.S., A.M. D'Gama, and C.A. Walsh, *Chapter 12 - Somatic Mosaicism and Neurological Diseases*, in *Genomics, Circuits, and Pathways in Clinical Neuropsychiatry*, T. Lehner, B.L. Miller, and M.W. State, Editors. 2016, Academic Press: San Diego. p. 179-199.
92. Kumar, K.R., M.J. Cowley, and R.L. Davis, *Next-Generation Sequencing and Emerging Technologies.* Semin Thromb Hemost, 2019. **45**(7): p. 661-673.
93. Barzon, L., et al., *Applications of next-generation sequencing technologies to diagnostic virology.* International journal of molecular sciences, 2011. **12**(11): p. 7861-7884.
94. Qin, L., et al., *Detection and Quantification of Mosaic Mutations in Disease Genes by Next-Generation Sequencing.* The Journal of Molecular Diagnostics, 2016. **18**(3): p. 446-453.
95. Maruyama, T., et al., *Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.* Nature Biotechnology, 2015. **33**(5): p. 538-542.

96. Kaneko, T., *Genome Editing in Mouse and Rat by Electroporation.* Methods Mol Biol, 2017. **1630**: p. 81-89.
97. Hsu, P.D., et al., *DNA targeting specificity of RNA-guided Cas9 nucleases.* Nature biotechnology, 2013. **31**(9): p. 827-832.
98. Xu, H., et al., *Sequence determinants of improved CRISPR sgRNA design.* Genome research, 2015. **25**(8): p. 1147-1157.
99. Doench, J.G., et al., *Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation.* Nature Biotechnology, 2014. **32**(12): p. 1262-1267.