Development and use of novel inducible Cas9 models to study gene function

By

Vignesh Krishnamoorthy

A thesis submitted to the Faculty of Graduate studies of the University of Manitoba in partial fulfilment of the requirements of the degree of

Master of Science

Department of Pharmacology and Therapeutics

University of Manitoba

Winnipeg, Manitoba

Table of contents

Abstract

Acknowledgements

List of tables

List of figures

List of abbreviations

Chapter 1: Introduction……...1

1.1 Mice in the realm of cancer……..1

1.2 Genetically engineered mouse models (GEMMs) for cancer……... …..2

1.3 Mouse embryonic stem cells based transgenesis…….....3

1.4 Mouse ES cell transgenesis toolbox…….....5

1.41 Cre/LoxP system and drug inducible system……5

1.42 Safe Harbour loci for gene targeting…….7

1.43 Recombinase mediated cassette exchange (RMCE)…….7

1.44 Gateway cloning technologies…….....8

1.45 Combination of Cre/LoxP system, drug inducible system and Gateway cloning Technologies……..10

1.5 CRISPR-Cas9 for gene modulation……13

1.6 Applications of CRISPR-Cas9 technology…….....15

1.61 CRISPR/Cas9 based genome wide screens……15

1.62 CRISPR/Cas9 system for general mouse modeling……16

1.7 Gibson's assembly……..18

1.8 Aims of the project…….....19

Chapter 2: Materials and methods

2.1: Mouse work and sarcoma cell line establishment……..20

2.2: Cell culture: Mouse embryonic stem cell culture

2.21 ES cell culture…….....21

2.22 Passaging of mES cells……...21

2.23 Cryopreservation of mES cells……21

2.24 Thawing frozen cells…… ..22

2.25 ES cell line establishment……...22

2.3 ES cell DNA isolation…….....22

2.4 *Trp53* targeting gRNA containing plasmid……22

2.5 Lentivirus Production…….....23

2.6 Quantitative Reverse transcription polymerase chain reaction……..23

2.7 Multi site Gateway cloning - (L-R reaction) …….....23

2.8 Targeting the wt/dCas9-MPH vectors into mES cells…….....24

2.9 Competent cell transformation……24

2.10 Plasmid DNA isolation…….24

2.11 Restriction digestion…….....25

2.12 PCR product isolation……...25

2.13 DNA Sequencing……..25

2.14 Fluorescence activated cell sorting (FACS)…….....29

2.15 X-gal staining of cells……...29

2.16 T7 Endonuclease assay…….....29

Chapter 3: Results

Part 1: 3.1: To utilize a doxycycline-inducible Cas9 mouse model and ES cell derivatives for tumor modeling and guide RNA validation respectively……30

3.1.1 Background……...30

3.1.2 Establishment and validation of mES cells from the Cas9 inducible mouse…….30

3.1.3 DOX inducible ES cell system as an *in vitro* platform to validate the function of a commercial guide RNAs……...36

3.1.4 Evaluation of the Cas9 inducible mouse model for tumor modeling……41

Part 2: 3.2 To develop a mouse line with the conditional and inducible (COIN) expression of wt/dCas9 and the MPH gene activation complex…….46

3.2.1 Background……46

3.2.2 Development and characterization of positive control mES cell lines……...50

3.2.2.1 Development of positive control mES cell lines for the COIN wtCas9-MPH system.50

3.2.2.2 Development of positive control LacZ mES cell lines for DOX induction experiments……..53

3.2.2.3 Flow cytometric characterization of the three positive control mES cell lines……56

3.2.2.4 Functional characterization of constitutive wtCas9-MPH expressing cells……..58

3.2.3 Generation and characterization of the conditional and inducible (COIN) wt/dCas9-MPH expressing mES cell line……...60

3.2.3.1 Development of in cis conditional and inducible (COIN) wt/dCas9-MPH expressing mES cell line at the R26 locus……..60

3.2.3.2 Functional characterization of the COIN wtCas9-MPH mES cells…….66

Chapter 4: Discussion…….68

Chapter 5 Conclusion and Future Directions……..……74

Abstract

CRISPR/Cas9 based mouse models are gaining traction for *in vivo* disease modeling. However the existing models do not allow both the spatial and temporal control of Cas9 expression which may lead to off-target effects. This thesis employed two different approaches to utilize/develop Cas9 models to replicate tumorigenesis with reduced off target effects. The first approach illustrated the utility of a drug inducible Cas9 model to recapitulate sporadic tumor development when Cas9 is induced along with the delivery of gRNAs targeting desired genes. In the second approach, we have attempted to develop a mouse model wherein both gene inactivation and gene activation could be achieved by creating an ES cell model with both Cas9 and a gene activation complex. ES cells expressing conditional and inducible variants of Cas9 and the gene activation complex has been created and initial data suggests that the system should be expressing Cas9 and the gene activation complex.

Acknowledgements

First and foremost, I would like to thank my supervisor Dr. Jody Haigh for his unconditional support and mentorship over the past 2.5 years and for being an amazing supervisor. I'm so thankful to him for providing me an opportunity to work in an amazing project in his lab. I learnt so much from him from the past 2.5 years right from designing good experiments to using proper positive controls to scientific writing. I whole heartedly thank him very much for his kindness and generosity in understanding my situation and helping me to graduate at the earliest for facilitating my move to another province.

Many thanks to my committee members, Dr. Geoff Hicks, Dr. Tamra Ogilvie and Dr. Sachin Katyal for their valuable time and support during the committee meets. I'm in particular very grateful to Dr. Ogilvie for her kind support during my PhD admission process.

My heart felt thanks to Katharina Haigh for being the super lab mom in ensuring the lab runs in a smooth manner. Many thanks for teaching me bacterial cloning, mouse ES cell culture and most importantly for making MEFs relentlessly. Also, thanks a lot for the delicious treats provided over the holidays.

I thank Aissa so much for his patience in teaching different experiments over the course of my masters. It was so wonderful to learn many intricate details of numerous molecular techniques from you. Many thanks for your kindness for being ever ready to help me with my experiments. It was so inspiring to see how you work in the lab.

Thank you, Andrew, for being one of the kindest persons I've ever seen, thank you for your help in the animal experiments.

I also thank Carlos, former member of the Haigh lab for being a good friend during his time in the lab. Thanks for your company during lunch. It was so nice to know about you and Chile.

Thanks to Tyler for being a warm and friendly colleague.

I thank the funding received from University of Manitoba Graduate fellowship, Research Manitoba, Cancer Care Manitoba and GETS fellowship during my masters.

Personally, I would like to thank my friends and family for their constant support. Many thanks to my parents for providing me the best education in India and all the necessary means that enabled me to pursue my studies in Canada.

List of tables

Table 1: List of DNA sequencing primers…….27

List of Figures

NO	List of Figures	PNO
1.1	Development of a transgenic mouse model through ES cell transgenesis	3
1.2	Cre/LoxP system	5
1.3	Tet on system	6
1.4	Trap mediated FlpE Recombinase Mediated Cassette Exchange Reaction (RMCE)	8
1.5	Gateway cloning	9
1.6	Detailed pictorial representation of generating and targeting the polycistronic conditional and inducible (COIN) cDNA vectors into docking mES cells	11,12,13
1.7	CRISPR- Cas9 technology for creating gene knock outs	14
1.8	CRISPR- Cas9 technology for endogenous gene expression modulation	15
3.1	Establishment of ES cell lines from Gt (ROSA)26Sortm1(rtTA*M2)Jae Col1a1tm1(tetO-cas9)Sho mice (iCas9 mouse) model using the 2i protocol	34
3.2	qPCR analysis illustrates Cas9 expression induced in the iCas9 mES cells only when treated with doxycycline	36
3.3	Lentiviral transduction of the plasmid containing a gRNA targeting *Trp53* into iCas9 mES cells	39
3.4	T7 endonuclease assay reveals the mutations created by Cas9-gRNA in the *Trp53* gene of the iCas9 mES cells transduced with a guide targeting *Trp53* and induced with DOX	41
3.5	iCas9 mice form tumors when injected with AAV9 cassette containing guides targeting *Trp53*, *Lkb1*, *Kras* and homology arms for introducing oncogenic Kras and induced with doxycycline	44
3.6	Hematoxylin and Eosin staining of the tumor tissues and the cell lines established from them reveals the herringbone morphology and fibroblastic phenotype suggesting that the tumor may be a fibrosarcoma	45
3.7	Sanger sequencing depicts the irregular reads in the *Lkb1* gene in a cell line established from the tumor compared to the wild type muscle cell line	46

3.8	Schematic overview of the development and utility of the conditional and inducible (COIN) Cas9-MPH mouse and its utility in simultaneous gene activation and gene inactivation	50
3.9	Generation of constitutive wtCas9-MPH mES cell line and inducible wtCas9-MPH mES cell line	53
3.10	Generation and validation of inducible LacZ expressing mES cell line	56
3.11	Flow cytometer profiling of reveals that large levels of constitutive wtCas9-MPH and DOX induced LacZ mES cells express EGFP	58
3.12	T7 endonuclease assay reveals the mutations created by Cas9-gRNA in the *Trp53* gene of the constitutive wtCas9-MPH mES cells transduced with a guide targeting *Trp53*	60
3.13	Detailed pictorial representation of generating and targeting the polycistronic conditional and inducible (COIN) wt/dCas9-MPH vectors into the docking mES cells	64
3.14	Conditional and inducible (COIN) wt/dCas9-MPH mES cell development	66
3.14	PCR analysis on putative Cre excised COIN wt/dCas9-MPH mES cell clones reveals the loss of 5’ integration band	68

List of Abbreviations

AAV: Adeno associated virus

Attp: attachment site

Bp: Base pairs

COIN: Conditional and inducible

ColA1: Collagen Type 1 Alpha 1 Chain

CRIPSR: **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **r**epeats

CMV: Cytomegalovirus

dCas9: Dead Cas9

DgRNA: Dead guide RNA

DMEM: Dulbecco's modified Eagle's medium

DOX: Doxycycline

DV: Destination vector

DPC: Day's post coitum

EFS: Elongation factor 1α short promoter

EGFP: Enhanced green fluorescent protein

FACS: Fluorescence activated cell sorting

FBS: Fetal bovine serum

FRT: Flp recognition site

Fw: Forward

G418: Geneticin

gRNA: Guide RNA

Ins: insulator core sequences

IP: Intraperitoneal injection

iPSCs: Induced pluripotent stem cell

IRES: Internal ribosome entry site

ITR: Inverted terminal repeats

KRAB: Kruppel associated box protein

LTR: Long terminal repeats

MEFs: Mouse embryonic fibroblasts

mES: Mouse embryonic stem cells

MitC: Mitomycin C

MPH: MS2-p65-HSF1

Mut: Mutant

Ori: Origin of replication

OSKM: Oct4, Sox-2, Klf4 and c-Myc

PBS: Phosphate buffered saline

PGK: Phosphoglycerate kinase

R Fw: ROSA Forward

R Rw: ROSA Reverse

Rluc: Reporter luciferase

RMCE: Recombinase mediated cassette exchange reaction

ROSA26: Reverse oriented splice acceptor clone 26

Rpm: Rotations per minute

(r)tTA: Reverse tetracycline trans activator

SA: splice acceptor

SAM complex: Synergistic activation mediator

SPH: Sun tag-p65-HSF1

STOP: Transcriptional STOP tPA sequence

Tet: Tetracycline

tetO/R: Tet-Operon

TRE: Tet responsive element

X gal: 5 bromo-4-Chloro-3-indoyl-β-D-galacto-pyranoside

Development and use of novel inducible Cas9 models to study gene function

Chapter 1: Introduction

1.1 Mice in the realm of cancer

Mus musculus or the common mouse is the most widely used mammalian organism in biomedical research and its use is not only limited to studying various disease processes but has also been invaluable in understanding numerous aspects of mammalian developmental genetics as a part of comparative biology[1]. Benefits of using the mouse as a model organism include its similarity to humans in terms of anatomy, physiology and genome (on average, the protein-coding regions of the mouse and human genomes are **85 percent identical)**, ease in maintaining large colonies and most importantly its genome can be readily modified. Gene function characterization in normal and disease states have been possible with the help of mouse models which have led to the better understanding of human diseases and designing novel therapeutics[2]. However, it should be noted that there are still some limitations pertaining to mouse models mainly owing to differences in species which affects the dynamics of cancer susceptibility, innate cellular signaling pathways and oncogenic transformation dynamics to name a few[3].

The mouse has been an integral part of modern cancer research since the mid 20th century. Cancer is a disease which occurs because of aberrant proliferation and survival of malignant cells. Cancer complexity arises through the extensive interactions between different cellular components and consists of many hallmarks such as oncogenic transformation, tumor metastasis and angiogenesis among others[4]. These processes can be best studied only *in vivo*. Various mouse models have identified several key mechanistic aspects of cancer. They are a well-established platform for validating the findings from human tumors and cell lines. This has substantially refined our understanding of the complex nature involved in the development and progression of cancer. Pre-clinical mouse models are essential in determining the safety, toxicity and efficacy of potential therapeutic drugs. They have helped to screen a myriad number of drugs in pre-clinical settings to assess both toxicity and well as efficacy[5,6]. There are different types of cancer models in mice and numerous approaches are used for modelling. Mouse models for cancer are broadly classified under two different categories: tumor xenograft models and genetically engineered modified mouse models (GEMMs)[5]. Xenograft models are generated by injecting either murine tumor cells or human tumor cells usually into severe immune compromised mice that allows the tumor cells to engraft in the animal without getting rejected[7]. These models are widely used to study the response of a therapeutic drug against the transplanted tumors. These models are also widely described as “animal cultures”[8]. The advent of recombinant DNA and

gene manipulation technologies in the latter half of 20th century led to the development of genetically engineered mouse models which were very beneficial for cancer studies[6].

1.2 Genetically engineered mouse models (GEMMs) of cancer

GEMMs are developed by manipulating the mouse genome either at the germline or at the somatic cell level. Manipulating the mouse genome has shed light on normal developmental processes and how when gene expression goes awry can lead to deregulated signaling which contributes to cancer[9]. The tumors initiated in GEMMs by introducing desired mutations in putative oncogenes and tumor suppressors closely resembles tumors from human counterparts and in many cases also mirrors tumor heterogeneity observed in humans[9,10]. GEMMs develop tumors amidst a healthy immune background and intact microenvironment and as a result of which the influence from these components on tumorigenesis can be studied in a precise manner[9]. Engineering desired cell specific changes in a mouse has been the holy grail in deciphering the intrinsic gene regulatory mechanisms responsible for maintenance of genome integrity and how any perturbations in these key events can lead to cellular transformation and tumour progression[6]. Creation of different mutations in a single mouse helps to understand the key interplay between different cell specific pathways responsible for driving the tumorigenesis and drug resistance which aids in the development of better therapeutics. These GEMMs have been a valuable platform for identifying the novel roles of different genes and also for characterizing the prospective tumor associated genes identified through sophisticated bioinformatics approaches such as RNA sequencing or single cell RNA sequencing[11].

The pioneering experiments of Drs Rudolf Janeisch and Betrice Mintz showed that microinjection of early-stage mouse embryo with simian virus 40 viral DNA (SV 40) can give rise to progeny containing SV40 in their DNA[12]. Although, the resulting progeny neither developed tumors even after 1 year of age nor were able to transmit the viral DNA to the next generation, this experiment paved the way for developing the first genetically engineered cancer mouse models or commonly dubbed as oncomouse[12]. Four independent groups in the early 1980s developed different variants of the oncomouse. Richard Palmiter and Ralph Brinster fused parts of the SV 40 T antigen DNA to a recombinant gene - MK (consisting of the mouse metallothionine gene promoter and thymidine kinase) and injected the SV-MK construct into the zygote which gave rise to transgenic mice carrying the SV-MK gene. This mouse and its offspring were predisposed to developing brain tumors. Philip Leder and Timothy Stewart fused the LTR (long terminal repeats) of MMTV (mouse mammary tumor virus) with *c-Myc* and injected them into zygotes and the resulting offspring were predisposed to developing mammary tumors. Erwin Wagner fused metallothionein promoter with the Fos gene fused with LTR

from a sarcoma virus, which upon injecting into zygotes gave rise to a transgenic colony that developed benign bone lesions but did not develop any fully transformed tumors. However, in the subsequent study, the same group fused MHC class 1 promoter with the Fos LTR which upon injection into zygotes gave rise to a colony of mice that developed bone tumors and was the first transgenic mouse model of osteosarcoma. Douglas Hanahan merged the rat insulin gene with parts of the SV40 viral genome and created the RIP-Tag which when introduced into the zygote resulted in the generation of a litter wherein most of them carried the gene. Furthermore, subsequent breeding of this mouse gave rise to transgenic offspring developing pancreatic tumors. Similarly, the same group injected repeats of the bovine papilloma virus into mouse embryos which gave rise to transgenic mice that developed skin cancer[12]. Thus, a new era in mammalian developmental genetics was born after the seminal discoveries of technologies leading to the creation of oncomouse. Later the isolation of mouse

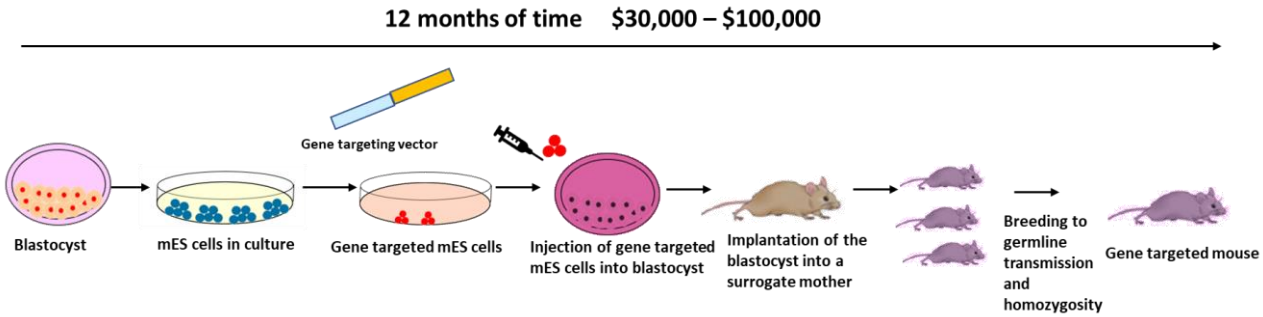


Figure 1.1: Development of a transgenic mouse model through ES cell transgenesis

A schematic illustrates the procedure for developing a transgenic mouse model wherein the mES cells are manipulated through gene targeting vectors which upon injecting into a host blastocyst and subsequent implantation into a surrogate mother gives rise to chimeric offspring. The chimeric mouse is then bred to homozygosity to create a null mutation or a knock out mouse.

embryonic stem cells (ES cells) and numerous gene editing tools augmented the development of GEMMs and led to the development of mouse models mimicking numerous aspects of human diseases[13].

1.3 Mouse embryonic stem cells based transgenesis

Mouse embryonic stem cells (mES) were first isolated by Martin Evans in the 1980s from the inner cell mass of the preimplantation (blastocyst stage) embryo. They are pluripotent in nature and have the ability to give rise to all the cells of the three germ layers (ectoderm, endoderm and mesoderm). This property was harnessed to demonstrate that ES cells can contribute to the adult tissues and also to the germline of chimeras when injected into a host blastocyst and implanted into a surrogate mother. Shortly thereafter, the ability to exchange an endogenous allele for another gene through homologous

recombination was achieved in mammalian cells *in vitro* conditions. This technique led to the generation of the Nobel winning gene targeting technology based on homologous recombination that enabled to perform site specific modifications in mES[14]. Gene targeting is a procedure wherein locus specific gene manipulations are introduced into the mES cells for creating GEMMs. It involves designing a plasmid vector containing a gene fragment for introducing the desired modification, a positive selectable marker for selecting the correctly targeted mES cells along with two homology arms for recombination into the targeted loci[15].

Desired genetic modifications for recapitulating various human disorders can be introduced into mES cells and the resulting mouse developed from these ES cells carry the genetic modifications in them that can be transmitted to the next generation. The whole process of generating a gene targeted mouse model usually takes about 6 months - 1year of time and it's also a labour intensive and a costly process (see **Figure 1.1**)[14,16]. A null allele or a knock out is usually created by disrupting crucial exons by replacing it with a drug resistant marker or a reporter cassette. Knock out approaches have greatly enhanced the understanding of the function of numerous developmental related genes and tumour suppressors. A knock in allele is created by incorporating a new sequence into the target allele[17]. Knock in approaches enabled a better understanding of the role of a mutated gene by introducing mutated alleles into the endogenous locus of numerous genes[17]. The introduction of newer gene targeting technologies hastened the efforts of developing mice models that closely resemble human disease counterparts.

The Cre/loxP system allowed for the creation of conditional knock outs and cell/tissue specific control of the genetic alteration. Drug inducible systems were also developed that allowed temporal control of gene expression by placing it under the control of a drug responsive promoter element. The introduction of reporter genes into any endogenous gene locus has enabled us to track the expression of the specific gene during *in situ* embryogenesis[14]. Technologies, like Gateway cloning, have enabled researchers to generate targeting vectors containing different gene fragments through a simple and efficient reaction and have accelerated the process of gene targeting[18]. Trap coupled recombinase mediated cassette exchange reactions (RMCE) have improved the speed of picking correctly targeted cells[19]. Developing complex transgenic mouse models solely based on the conventional gene targeting approaches such as Cre/loxP system and/or drug inducible system is time consuming and cost ineffective[16]. Multiple crosses have to be performed to generate triple transgenics which also involves generation of animals with unwanted alleles. However, the discovery of CRISPR/Cas9 technology and its subsequent utility in mammalian genome editing has revolutionized the ability to model human

diseases in mice with reduced breeding strategies and less expense[20]. Highlighted below are the annals of mouse ES based transgenesis and the tools involved in it that have been utilized in this thesis.

1.4 Mouse ES cell transgenesis tool box

1.41 Cre/loxP system and drug inducible system

The development of the Cre/loxP system allowed to conditionally delete the gene of interest and revolutionized genome editing approaches as many early embryonic lethal phenotypes associated with global gene deletion were avoided[16]. The Cre recombinase is a 35 kDa site-specific recombinase

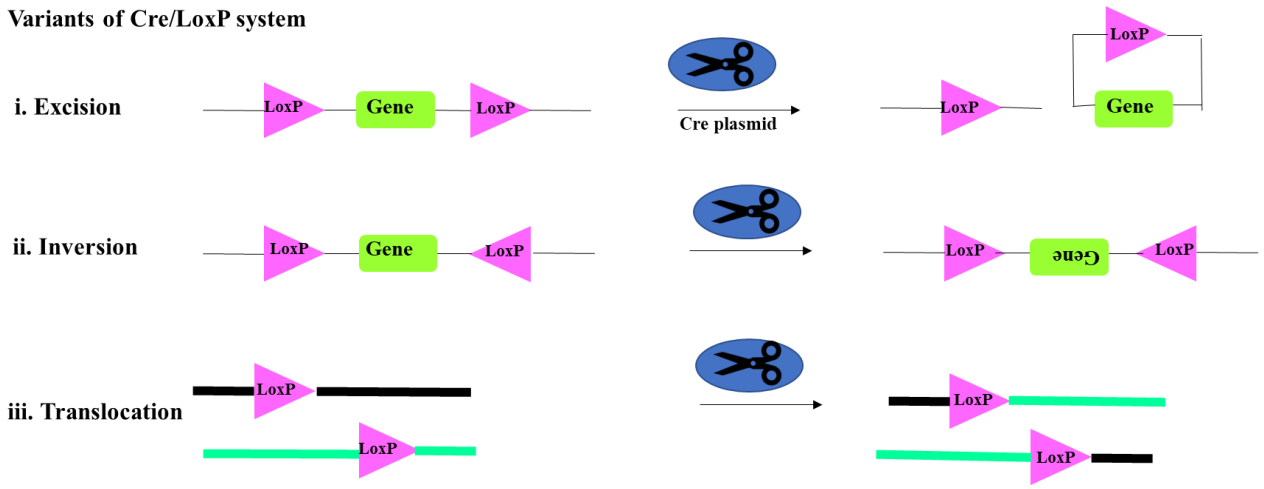


Figure 1.2: Cre/LoxP system

A schematic representing the different types of mutations that could be engineered with the Cre/LoxP system just by altering the orientation of the LoxP sites flanking the desired gene of interest. (Figure adapted from addgene https://www.addgene.org/collections/cre-lox/)

discovered from the bacteriophage P1 and orchestrates recombination between a pair of loxP sites. The loxP site (34 base pairs) consists of two 13 bp inverted repeat sequences that are separated from each other by 8 bp sequence that determines the position of the loxP site. The Cre recombinase removes any DNA sequence present between a pair of loxP sequences if they are in the same orientation and causes an inversion if the loxP sequences are present in the opposite orientation. If the loxP sites are in two different chromosomes, then the Cre enzyme can perform translocation of the sequences between chromosomes but very infrequently (see **Figure 1.2**)[16]. The incorporation of this system for gene targeting in ES cells has helped to investigate the functions of numerous genes which were unable to be deciphered from original gene targeting due to embryonic lethality or developmental arrest observed when some genes are deleted constitutively during development.

The *in vivo* Cre/loxP system typically relies *on* two different mouse lines. One strain would be expressing Cre enzyme from a promoter that is unique to a specific cell lineage or a tissue. The other strain would have loxP sites flanking the gene of interest to be deleted (often referred to as floxed or conditional allele). Upon breeding the cell/tissue specific Cre line with the conditional mouse line, Cre would be expressed only in the specific tissue of interest and the corresponding gene would be deleted only in that specific tissue. This approach allows researchers to delete the gene of interest in a conditional manner. Presently, there exist many cell/tissue-specific Cre lines (see https://www.jax.org/research-and-faculty/resources/cre-repository/characterized-cre-lines-jax-cre-resource). As well, several international consortia have created many conditional ES cell lines that researchers can obtain for studying their specific gene of interest in a cell/tissue restricted manner (for example see https://www.eummcr.org/products/es-cells). Drug inducible Cre/loxP systems have also been developed wherein the Cre enzyme is modified (e.g., through fusion with parts of the oestrogen-receptor) and gets activated only with the addition of a drug such as tamoxifen (an oestrogen analogue). This enables the deletion of a particular gene in a specific cell/tissue at desired time point during development or in the adult[21]. One drawback to the Cre/loxP system is that once the Cre is active then that cell and all progeny of that cell will have deleted the gene of interest.

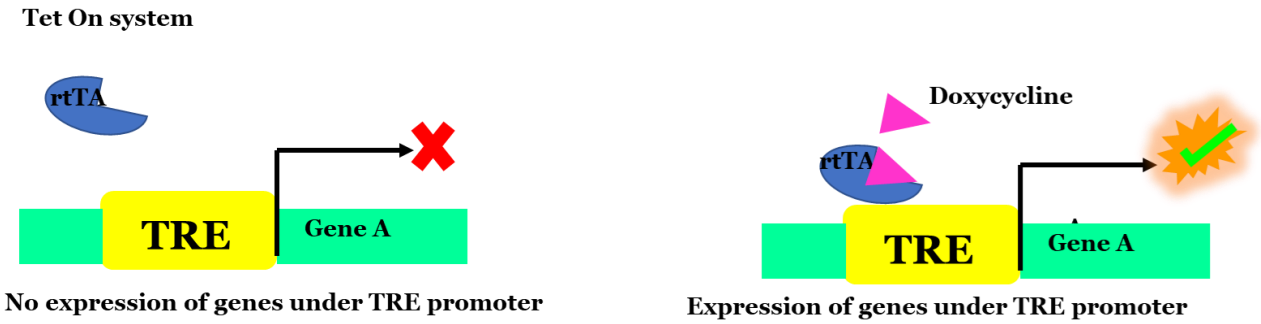


Figure 1.3: Tet on system

A schematic illustrating the working of the tetracycline on system. The system consists of tetracycline responsive element (TRE promoter) and rtTA (reverse tetracycline trans activator). When the system is stimulated with doxycycline, DOX binds with rtTA and drives the expression of genes under the control of TRE promoter. (Figure adapted from addgene:// www.addgene.org/collections/tetracycline)

To overcome these issues, additional drug inducible systems have been developed that allow researchers to turn on and turn off a gene's expression at desired time intervals in the presence/absence of a drug. The tetracycline based on/off system is one of the most frequently used systems for modulating gene expression owing to its tight temporal control over transgene expression[21,22]. It is

originally based on the E. coli's Tet operon and is adapted for use in mammalian systems including humans cell lines. Two different variations have been created and they are the Tet on and Tet off system. Tet on system is the most exploited one and is currently used in this thesis.

The system consists of reverse tetracycline controlled trans activator (rtTA) (a transcription activation domain from virus that is structurally modified) and the *tet*-responsive element (TRE: tet(o)-CMV_{min}) which contains multimerized rtTA binding sites upstream of a minimal CMV promoter. Upon addition of doxycycline (DOX), a tetracycline analogue, rtTA binds to DOX and this complex binds to TRE promoter and drives the expression of genes downstream of it (see **Figure 1.3**)[21,22].

1.42 Safe Harbour locus for gene targeting

Researchers have used the ROSA 26 (R26) as a safe harbour locus for introducing the desired gene targeting vector containing transgenes of interest into the mES cells. It is formally denoted as Gt(ROSA) 26 Sor present on chromosome 6 and R26 targeted ES cells allows to achieve stable and ubiquitous gene expression in all the cells resulting from the transgenic animals[23]. The R26 locus was discovered through genetic screens for identifying insertional mutagenesis wherein a reporter gene encoding for promoter less beta galactosidase and neomycin resistance was introduced into mES cells through viral infection. The reporter activation was detected based upon the insertion of the reporter into an endogenous locus driving its expression and this locus was named as the ROSA26 locus. Targeting genes in a constitutive or conditional manner into R26 locus does not affect the developmental viability of the mES cells as the process does not disrupt the function of the endogenous genes at any levels[23].

1.43 Recombinase mediated cassette exchange (RMCE)

Traditional methods of site-specific gene targeting utilizing homologous recombination are inefficient and time consuming with the frequency of homologous recombination being very low and thus a large number of ES cell colonies have to be picked in order to identify the correctly targeted cells. This issue was circumvented by developing a site-specific gene targeting technology known as **r**ecombinase **m**ediated **c**assette **e**xchange (**RMCE**) reaction which enable the researchers to insert a desired vector sequence at a specific locus in an efficient manner[19]. RMCE is a relatively simple process by which efficient gene targeting to the desired locus can be achieved by exchanging a previously existing cassette for a new cassette (which is the gene targeting vector) carrying the genetic elements of interest[24]. Typically, a gene targeting vector is flanked by a pair of site-specific recombination sites and the target genomic locus in the targeting cells of interest is also flanked by corresponding pairs of recombination sites. Upon co delivery of the gene targeting vector into the desired cells along with the

specific recombinase allows to exchange the gene targeting vector in place of a resident cassette that is present between the two site specific recombination sites in the genomic locus of the target cells.

A trap coupled RMCE based targeting approach was developed wherein the gene targeting vector has a promoter for a drug resistant gene and the target loci in the mES cells has a promoter less drug resistant gene (see **Figure 1.4**). Correct targeting, confers drug resistance in the mES cells and selection of positive clones. We employ a FlpE recombinase-based reaction wherein the recombinase exchanges any sequence present between two corresponding pairs of FlpE recombination sites (Frt)[19].

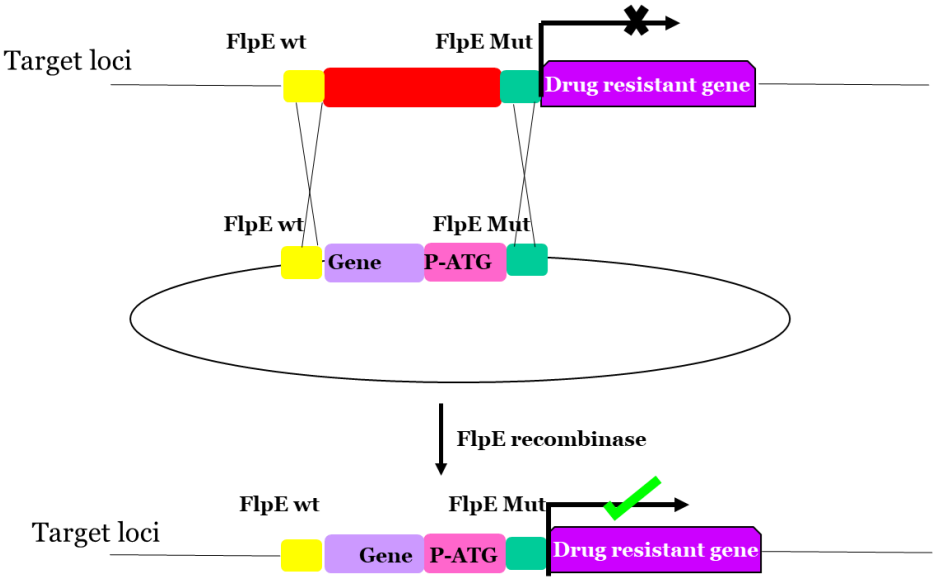


Figure 1.4: Trap mediated FlpE Recombinase Mediated Cassette Exchange Reaction (RMCE)

The schematic illustrates how the FlpE recombinase exchanges any sequences present between a pair of corresponding heterospecific Frt sites and the trap mediated approach allows to restore the promoter (P-ATG) for the drug resistance gene only upon correct targeting.

1.44 Gateway cloning technologies

Conventional gene targeting cloning procedures involving extensive use of restriction enzymes and ligases and can become cumbersome when multiple fragments are involved or there is a lack of appropriate restriction sites. Gateway cloning™ is a superior cloning system that was developed to ease the complications associated with traditional restriction enzyme-based directional and blunt-end cloning techniques. The cloning methodology employs precise integration and excision reactions that are a part of bacteriophage lambda's infectivity cycle into E. Coli[18].

The reaction involves three different steps, the first one entails incorporation of specific attachment (recombination) sequences known as attB1 and attB2 flanking both the sides of the gene of interest. This may be done through PCR or restriction digestion. The second step involves cloning the gene of

interest flanked with attB sites into a special Gateway donor plasmid containing compatible attP recombination sites that flanks the CCDB gene. This reaction is catalysed with the help of an enzyme known as BP clonase that contains the phage integrase and *E.Coli* integration Host factor (IHT). Here, the attB and attP sequences recombine and an entry clone containing attL flanking the gene of interest is created. The entry clone containing attL sites is then cloned into a destination vector containing CCDB gene flanked by attR sites on and other additional features such as drug resistance genes (see **Figure 1.5**). The cloning is mediated by a LR clonase enzyme that consists of phage recombination proteins Int, excisionase and IHT which exchanges the CCDB gene for the incoming gene of interest present in the entry clone. The end product of this cloning is the generation of an expression plasmid containing the gene of interest flanked by attB sites and another plasmid containing CCDB gene flanked by attP sites. The entry clone and the destination vector contain different antibiotic resistance markers that would help to select the correctly targeted clones after transformation. CCDB is a cell death gene that encodes a toxin that targets the DNA gyrase and as a result, *E.Coli* that harbour the CCDB gene will be eliminated and ensures that only the clones that contain the B-P/ L-R reaction product survive. The control of cell death gene (CCDB) gene encodes for a toxin that targets the DNA gyrase and as a result of which the E.Coli that harbours the CCDB gene would die. Prior to the B-P/L-R reactions, vectors that contain the CCDB gene are grown in CCDB resistant E.Coli strains.

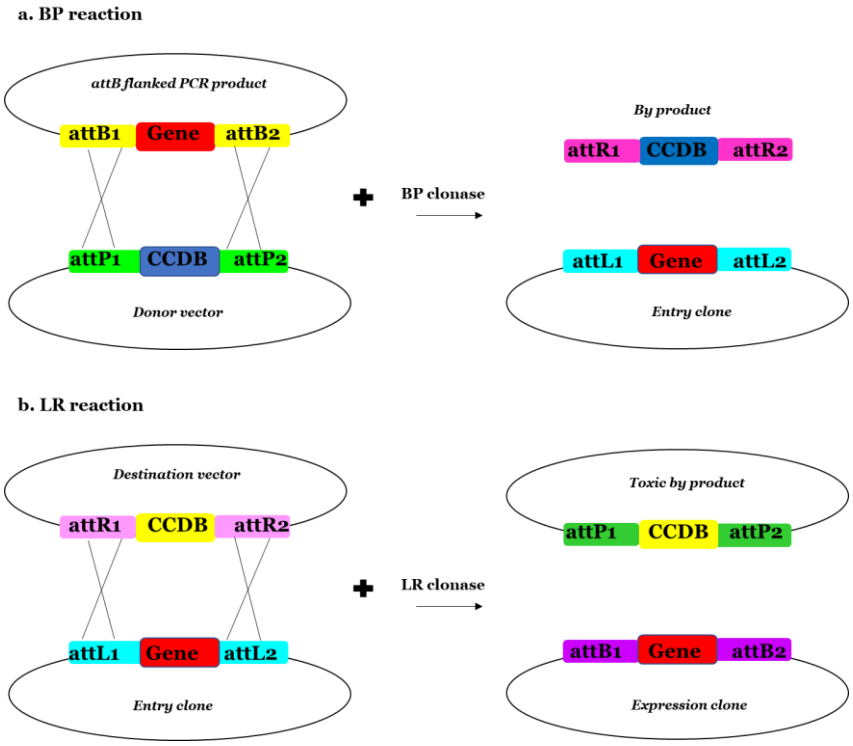


Figure 1.5: Gateway cloning

A schematic depicting the process of B-P (a) and L-R (b) reactions that constitute site specific recombination between different pairs of plasmids. (Figure adapted from addgene: https://blog.addgene.org/plasmids-101-gateway-cloning)

The main advantage of this system is that the entry vectors can be cloned into different destination vectors containing different features. Moreover, multiple pEntry vectors can be cloned into one Destination vector for complex gene manipulation strategies to be performed. Numerous Entry clone libraries suitable for a myriad number of appropriate destination vectors have been created and are available for use from commercial sources[18]. Within, this thesis several pEntry and Destination vectors were used for the creation of targeting vectors for RMCE reactions for transgene insertion into the Rosa26-docking site.

1.45 Combination of Cre/LoxP system, drug inducible system and Gateway cloning technologies

The Haigh lab has previously developed efficient conditional and inducible vector technologies for targeting genes to the safe harbor R26 locus present in mES cells[25,26,27]. Using the same technology, they also generated a novel mouse model for deriving induced Pluripotent Stem Cells (iPSCs) from adult cells from any of the three germ layers just by adding the DOX drug that turns on OSKM (Oct4, Sox-2, Klf4, c-Myc)[28] Yamanaka reprogramming factors. These technologies were built by utilizing the Gateway cloning system to combine the Cre/loxP and the DOX/rtTA tet(o)-inducible systems which conferred conditional and inducible expression of cDNA transgenes from the R26 locus of the mouse ES cells **(**see **Figure 1.6*)*** and in the somatic cells obtained from the adult mice[27].

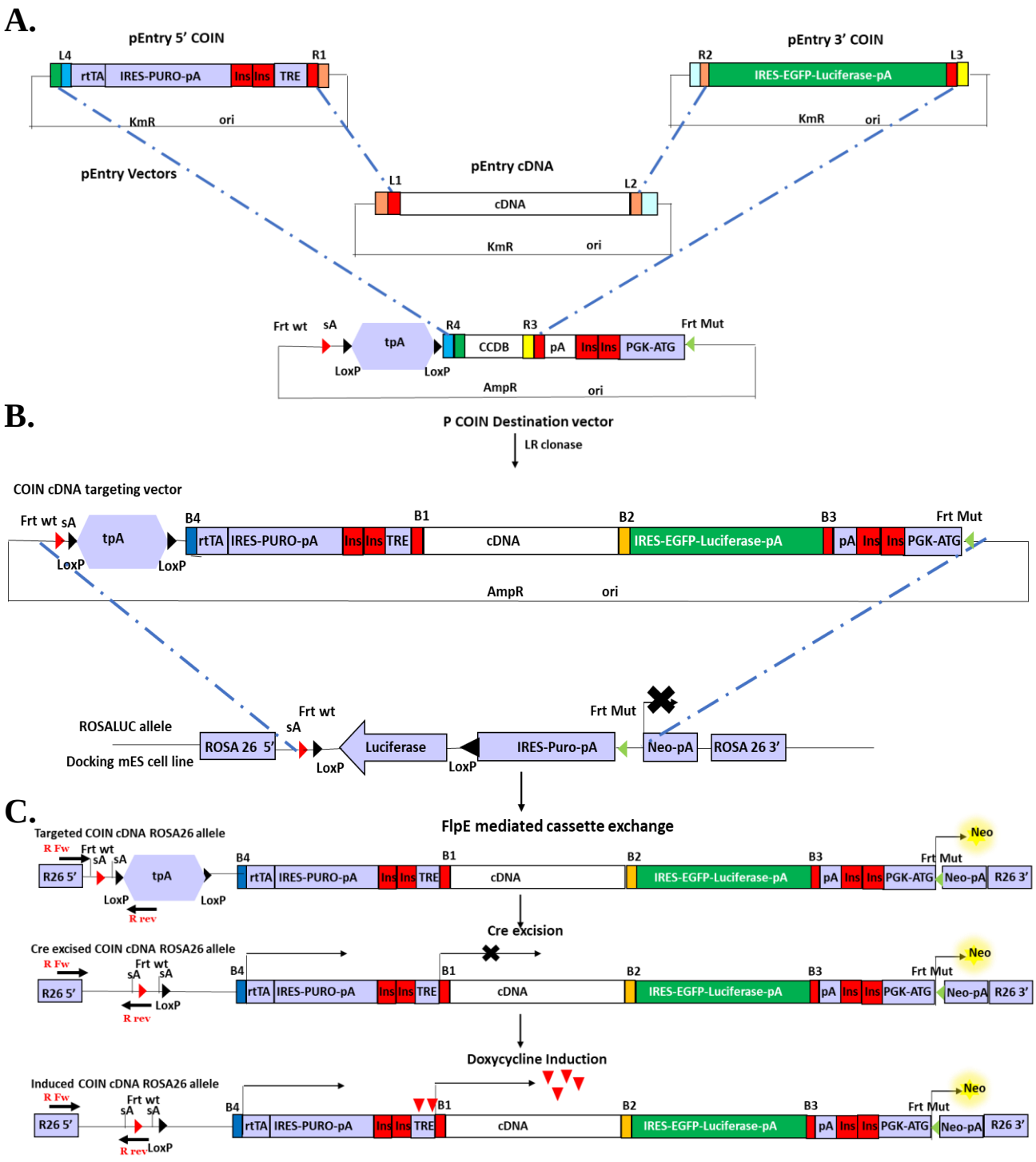

A.
pEntry 5' COIN
L4
rtTA
IRES-PURO-pA
Ins Ins
TRE
R1
KmR
ori
pEntry 3' COIN
R2
IRES-EGFP-Luciferase-pA
L3
KmR
ori
pEntry Vectors
pEntry cDNA
L1
cDNA
L2
KmR
ori
Frt wt
sA
tpA
LoxP
LoxP
R4
CCDB
R3
pA
Ins Ins
PGK-ATG
Frt Mut
AmpR
ori
B.
P COIN Destination vector
LR clonase
COIN cDNA targeting vector
Frt wt sA
tpA
LoxP
LoxP
B4
rtTA
IRES-PURO-pA
Ins Ins
TRE
B1
cDNA
B2
IRES-EGFP-Luciferase-pA
B3
pA
Ins Ins
PGK-ATG
Frt Mut
AmpR
ori
ROSALUC allele
Docking mES cell line
ROSA 26 5'
Frt wt
sA
LoxP
Luciferase
LoxP
IRES-Puro-pA
Frt Mut
Neo-pA
ROSA 26 3'
C.
FlpE mediated cassette exchange
Targeted COIN cDNA ROSA26 allele
R Fw
Frt wt
sA sA
R26 5'
tpA
LoxP
LoxP
R rev
B4
rtTA
IRES-PURO-pA
Ins Ins
TRE
B1
cDNA
B2
IRES-EGFP-Luciferase-pA
B3
pA
Ins Ins
PGK-ATG
Frt Mut
Neo
Neo-pA
R26 3'
Cre excision
Cre excised COIN cDNA ROSA26 allele
R Fw
Frt wt
sA sA
R26 5'
LoxP
R rev
Doxycycline Induction
Induced COIN cDNA ROSA26 allele
R Fw
Frt wt
sA sA
R26 5'
R rev
LoxP

Figure 1.6: Detailed pictorial representation of generating and targeting the polycistronic conditional and inducible (COIN) vectors into the docking mES cells

A. The three multi-site Gateway cloning compatible plasmid entry vectors and the destination vector were combined through LR reaction to generate the *in cis* conditional and inducible (COIN) cDNA gene targeting vectors. **B.** The COIN vector was then co transfected with a FlpE plasmid into the R26 locus of the GR4OSALUC cell line through trap mediated RMCE reaction which restores the neomycin resistance (Neo^R). **C.** Cre mediated removal of the STOP cassette results in the removal of the 3X pA-tpA+ STOP cassette and allows the expression of revere tetracycline controlled transactivator (rtTA) and the puromycin resistance gene as a bi-cistronic message from the endogenous R26 promoter. Inducible expression of cDNA can be achieved by administration with doxycycline which can also be monitored by the expression of EGFP. Correctly recombined clones could be screened by PCR (R F_w and R R_{Rev}) that confirms the amplification at 560 bp. (L1-4, R1-4, B1-4: Gateway compatible cloning sites, CCDB: control of cell death gene, SA: splice acceptor, IRES: internal ribosome entry sequence, pA: Polyadenylation sequence, Ins: Chicken beta globin insulator core sequence, TRE: Tet responsive element)

1.5 CRISPR-Cas9 for gene modulation

Advancements in bacteriology led to the discovery of CRISPR/Cas system. A popular variant is the CRISPR/Cas9 system and its use stretches from microbial technology to food processing to gene therapy. **CRISPR** stands for **c**lustered **r**egularly **i**nterspaced **s**hort **p**alindromic **r**epeats and was originally discovered as a part of prokaryotic antiviral defense mechanism, but later it was also modified to edit the mammalian genome. The CRISPR-Cas9 genome editing technology consists of the SpCas9 enzyme (also known as Cas9), a DNA endonuclease obtained from *Streptococcus pyogenes* and is dependent on a single guide RNA (sgRNA) for performing genomic site specific cleavage[30,31]. The single guide RNA (sgRNA) as the name implies guides Cas9 and its associated DNA cleaving (endonuclease) activity to the target sequence which is complementary to it and located upstream of the three nucleotide **p**rotospacer **a**djacent **m**otif (PAM) sequence. Cas9 has two domains for creating double stranded breaks (DSB) in the target DNA sequence and they are the HNH and RuVC like domain which cleave the target DNA sequence in the complementary and non-complementary DNA strands respectively (see **Figure 1.7**)[31]. Once the DSB occurs, then two different kinds of repair pathways can happen to mend the DNA break: the NHEJ - Non homologous end joining process and the HDR - Homology directed repair process. NHEJ usually results in large insertions or deletions – InDel mutations that can create knockouts. HDR is by far less frequent but can be used for precise gene editing if double stranded DNA targeting constructs or single stranded oligonucleotides are given as a repair template[32]. Many adaptations have been made to the classic CRISPR-Cas9 system through which activation/repression of genes can be accomplished with a high degree of accuracy without deleting the gene of interest. A refined version is the dCas9, a mutant form with altered DNA cleaving domains. dCas9 can still bind to the desired DNA sequence directed by the gRNA but cannot create double stranded breaks. Fusion of dCas9 with various gene regulatory

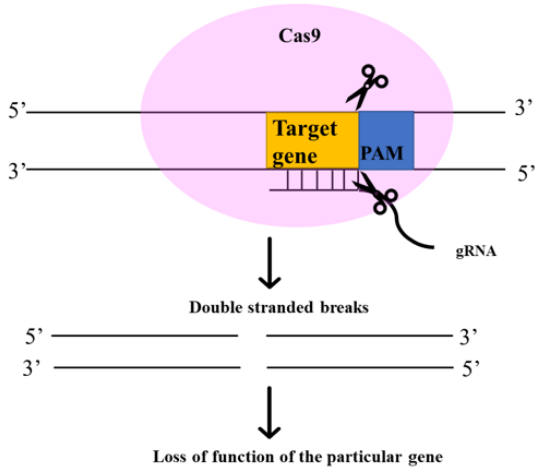


Figure 1.7: CRISPR- Cas9 technology for creating gene knock outs

A schematic illustrating how Cas9 binds to the target genomic sequence complementary to the guide RNA and creates a double stranded break which typically results in deletions or insertions.

domains enable fine tuning of endogenous gene expression in a precise manner[33,34]. Some of the regulatory domains that are utilized are repeats of VP16, VPR, MPH, Sun Tag based system and Com-KRAB (see **Figure 1.8**). VP16 is a DNA binding transcription activator obtained from the herpes simplex virus[35]. Multimerization of the VP16 domain can increase the potency of gene activation[36]. VP64, a multimer of VP16 has been reported to improve euchromatin chromatin signatures namely H3K27ac and H3K4me3[37] at the target chromatin loci[38]. VPR (VP64-p65-RTa) is a tri fusion protein designed by conjugating VP64 with activation domain of p65 (obtained from NF-kB[39]) and RTa (obtained from gamma Herpesviridae virus family[40]). It has been reported that merging of p65 and RTa robustly enhance the gene activation potential of VP64[41].

The MPH gene activation complex is designed by merging p65 and activation domain of human heat shock factor 1 (HSF1) and is held together by MS2-bacteriphage coat proteins. Guide RNAs harbouring MS2 hairpin aptamers are used as a scaffold to which MPH-wt/dCas9 complex can bind and activate the target genes. The combination of dCas9-VP64-MPH-MS2-gRNAs is a strong transcription activator known as synergistic activation mediator or SAM complex[42]. SPH activation domain (Sun tag-p65-HSF1) consists of Sun tag, a protein scaffold which holds p65 and HSF1 activation domain[43]. Reports have shown that targeting promoters of desired genes with multiple guides rather than single guide increase the efficacy of gene activation[44,45,46]. Com-KRAB is a gene repressor consisting of RNA binding protein (Com)[47] and Kruppel associated box protein (KRAB), a repression zinc finger[48] that disrupts transcription. Guides containing Com binding aptamers are used for recruiting Com-KRAB.

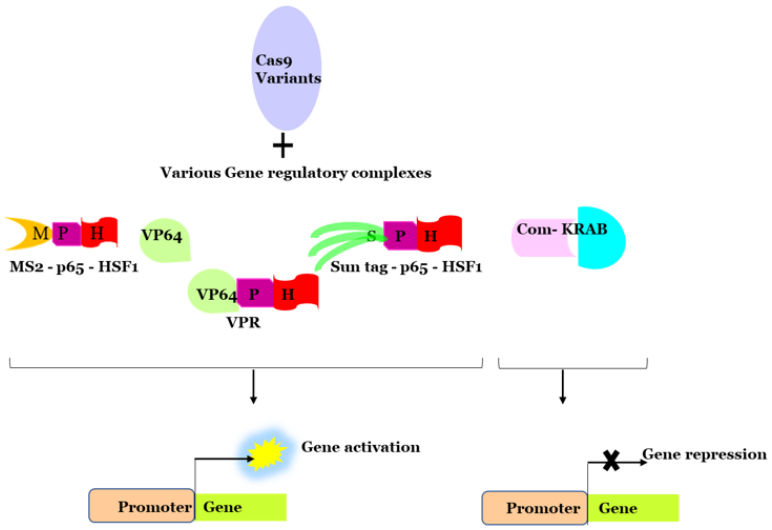


Figure 1.8: CRISPR- Cas9 technology for endogenous gene expression modulation

A schematic illustrating how different Cas9 variants conjugated with gene modulatory fusion proteins can result in gene activation or repression by binding to transcription start site.

Catalytically active Cas9 conjugated with gene modulating domains but guided by a shorter version of guide RNA known as dead guide RNA or DgRNA is an interesting alternative to dCas9 based strategies. DgRNA is created by removing few nucleotides from the normal guide which can still steer Cas9 towards the desired genomic location but at the same time prevents it from creating a proper DNA cleavage conformation[49,50]. Although clear differences between various CRISPR-Cas9 based gene modulator complexes are yet to be deciphered, it is generally observed that the gene activation potential increases with the addition of activation domains[38,41]. Cas9 variants conjugated with gene regulatory domains along with different guides are usually delivered through lentivirus for long term transduction[51]. Plasmid[52] and baculoviral vectors[53,54] have also shown high efficacy in delivering the Cas9 based gene regulatory components into the target cells. Cas9 variants conjugated with a myriad number of fusion domains can regulate the expression of endogenous genes and alter their epigenetic landscape[34]. This gene modulatory feature of Cas9 is becoming immensely popular in genome wide CRISPR screens, altering cellular identity of different cells and also for developing a myriad of *in vivo* disease models. A review is published in ReGen Open journal concerning the uses of CRISPR based technologies in regenerative medicine. (Krishnamoorthy and Haigh, CRISPR-Cas9 – The potential holy grail to generate biomedically relevant cells through cell fate engineering) Link to the article.

Considering my project's involvement in tumorigenesis modeling, applications of CRISPR/Cas9 towards the development of new cancer models is summarized below.

1.6 Applications of CRISPR-Cas9 technology

1.61 CRISPR/Cas9 based genome wide screens

Genetic screens have been an essential aspect in identifying the relationship between the changes at molecular level to the observed effect at the phenotype level. The ability of CRISPR/Cas9 technology to create changes in genome that can be performed in a quick and efficient manner has demonstrated its utility in large scale *in vitro* or *in vivo* genetic screens. CRISPR screens could be broadly categorised into knock out, knock down or activation screens. CRISPR screens involve transducing the desired cells usually expressing Cas9 with a viral library containing guides targeting a large repertoire of different genes with low multiplicity of infection such that only one sgRNA integrates into a cell. The transduced cells are subjected to a screening pressure such as drug selection wherein the relationship of the corresponding gene of interest to the selection is determined from the surviving cells that are either enriched or depleted for the corresponding sgRNA. All these changes in the sgRNA levels are identified by next generation sequencing technologies. CRISPR/Cas9 based *in vitro* knock out screens have been extensively used to identify the genes essential for the cancer cell growth, survival, metastasis, drug resistance and to decipher various aspects of tumor heterogeneity[55]. This kind of screening is hastening the process of developing a myriad of novel anti-cancer drugs against important targets. Furthermore, delivering pooled sgRNA library into cancer cell lines and subsequent

implantation into mice also identified novel drug targets for various cancer subtypes. Advancing a step further, direct *in vivo* CRISPR screens are also becoming popular as sgRNA libraries could be directly injected into a mouse expressing Cas9. The subsequent tumor development occurring in the occurring mouse allows to perform patho-physiologically more relevant screens compared to transplantation screens[55]. For example, Loganathan et al., utilized an *in vivo* tumor screens to identify the crucial role of Notch inactivation mutations as a hallmark for head and neck cancers[56]. The system is also being adapted to human organoids which would help discover novel hits and advance drug research in a more high-throughput manner, while at the same time preserving aspects of tumour cell heterogeneity[55].

1.62 CRISPR/Cas9 system for general mouse modeling

CRISPR/Cas9 technology has been widely used to create germline and somatic mice models. Germline knock out models are created by microinjection of the zygote or ES cells with guides RNAs and Cas9 targeting the desired gene of interest provided they do not initiate developmental arrest. CRISPR/Cas9 can knock out 5 different genes simultaneously in mouse ES cells and can be used to create double gene mutations harboring mouse in a single step by just delivering cas9 and the guides RNAs targeting the desired genes of interest into the zygote[57]. Mice harboring reporters have been developed by co injecting zygote with Cas9, guides targeting the desired gene's STOP codon or the UTR region along with a homology arm containing the reporter's sequence for knock in[58].

Somatic mouse models are developed by directly modifying the cells of the adult mouse. This approach is increasingly becoming popular due to its ability to model tumorigenesis in an efficient manner. Chromosomal structural abnormalities such as intrachromosomal inversions and deletions have been created in mouse livers by delivering guides and Cas9 targeting two regions in the same chromosome[59,60]. Numerous somatic mouse models were initially generated by co delivering Cas9 and the guides together into a mouse using plasmid and viral based vectors with the latter being the most dominant mode of delivery. This approach was widely used to model different pathological phenotypes associated with the liver owing to its easy accessibility by hydrodynamic tail vein injections. However, even with the viral vector-based delivery, it is quite cumbersome owing to the size of Cas9 protein[59,60]. This issue was greatly circumvented by developing Cas9 knock in mouse models, a type of Cas9 based mouse model that is of interest to this thesis. This type of mouse modeling approach involves introducing the conditional or inducible variants of Cas9 at the level of ES cells itself by knocking in the Cas9 at a ubiquitous locus like ROSA26 or Col1a1 locus. This allows to restrict the expression of Cas9 to a specific tissue of interest or induce its expression at desired time periods. This feature has

also abrogated the immunogenicity issues associated with the delivery of Cas9 proteins[20]. In this section, I will be discussing some of the Cas9 knock in somatic mouse models and how they have been applied in therapeutic settings or for disease modeling.

Cre dependent Cas9 knock in mouse.

This mouse model was created by knocking Cas9 and EGFP in the ROSA26 locus of mES cells that is blocked by a STOP cassette flanked by loxP sites. The mouse model developed from this gene targeted mES cells was successfully used to create a model of lung cancer[20]. They utilized an AAV vector that contains *rennila* luciferase, Cre recombinase, donor template for $Kras^{G12D}$, guides targeting *Kras, Trp53* and *Lkb1* and packaged it into an AAV9 virus. Upon delivering the AAV9 virus into lungs, they were able to model lung cancer recapitulating the complex heterogeneity observed in tumors derived from human patients. Moreover, the interactions between different mutated genes in promoting tumor development were also well documented through this model by tracking the Cas9 induced indels through deep sequencing[20]. This is a novel aspect of mutation tracking that was only possible through Cas9 based disease modelling and thereby reenforcing the superiority of this approach[20]. However, this model lacks the temporal control over the expression of Cas9 which may lead to off-target effects and thus may not be the most desired model for precise genetic manipulations.

dCas9-SPH knock in mouse

Zhou and colleagues, developed a tissue specific transgenic mouse line expressing dCas9 and SPH gene activation complex (Sun Tag-p65-HSF1) only in astrocytes. Delivering guides targeting three neural factors *Ascl1, Neurod1* and *Neurog2* into one side of the midbrain transdifferentiated astrocytes into neurons compared to the other part of the mid brain that received only control constructs[43]. This study illustrated the functionality of the dCas9 and the SPH complex that is embedded in the mouse and thwarts the need to deliver the huge dCas9 and the gene activation complex[43].

Cas9 knock in mouse

Belmonte's group[61] delivered the DgRNA-MPH gene activation complex for activating the expression of therapeutically relevant genes in a Cas9 knock in models of kidney disease and muscular dystrophy which alleviated these mouse models of disease greatly from the severity of these phenotypes[62]. Furthermore, they also illustrated that delivering DgRNAs and MPH complex targeting the promoter of the *Pdx1* gene in liver cells of this Cas9 knock in mouse with Type 1 diabetes transdifferentiated liver cells into functional pancreatic beta cells suggesting the stalwart abilities of this system to upregulate the activity of endogenous genes[62].

dCas9-MPH knock in mouse

Recently, Hunt et al., [63] developed a gene activation mouse model by targeting a conditional dCas9-MPH allele to the *R26* locus of mES cells and developing a mouse model from it. This was the first study to demonstrate the powerful aspect of the dCas9-MPH based endogenous gene modulation for disease modeling. This mouse line was effectively shown to model hypercholesteremia by upregulating the activity of endogenous *Pcsk9* which brought the cholesterol levels to those seen in human patients with high cholesterol. Similarly upregulating the activity of endogenous LDLR in high fat dCas9-MPH mice reduced the levels of circulating cholesterol and high-density lipoprotein[63].

These highlighted mouse models represent powerful tools for modeling diseases. However, they still lack either spatial (tissue specific) or temporal (inducible) control over the expression of Cas9 and or the gene regulatory complex. These two controlling elements are an essential part of disease modeling as they reduce the off-target effects associated with constitutive Cas9 expression. Moreover, for modeling complex *in vivo* gene interactions associated with tumor biology, simultaneous gene activation and gene inactivation needs to be done. Also, for understanding crucial cancer hallmarks such as oncogene addition[64] where the tumor may be solely dependent on the expression of a mutated oncogene, it is essential to have a system wherein the expression of an oncogene could be dampened after the tumor initiation. Although the dCas9 - gene activation complex containing mouse models allows to upregulate the expression of endogenous genes, they could not be used to perform gene inactivation owing to the inability of dCas9 to create double stranded breaks. Thus, there exists a need to develop *in vivo* models that allow simultaneous gene activation and inactivation.

1.7 Gibson's assembly

Gibson's assembly is an enzymatic reaction through which different gene fragments from different species can be cloned together through a simple thermocycler-based reaction provided they share some overlapping sequences. This reaction allows researchers to synthetize numerous large DNA fragments that were previously arduous to clone and assemble. The enzymes involved in the reaction are 5' exonuclease, DNA polymerase and a DNA ligase mixture which are available as a single reagent mix. Though they are present in a single reagent mix, the enzymes do not engage with each other. The exonuclease first acts and removes the overhangs from the 5' ends of both the DNA fragments, then the polymerase anneals the gaps between the two complementary molecules. Finally, the ligase closes the nick between the annealed DNA fragments. The exonuclease is sensitive to heat and its activity gets halted during 50ºC incubation[65]. Within this thesis Gibson Assembly was used to generated fusions between variants of Cas9 and the MPH transactivation complex.

1.8 Aims of the project

The overall aims of my project are to overcome some of the issues associated with the existing Cas9 based knock in mouse models for tumor modeling.

In the first aim, I aim to evaluate the capability of an existing mouse model that expresses Cas9 in a drug inducible manner for modeling sporadic sarcoma genesis and also to utilize ES cells established from this mouse as an *in vitro* platform for validating the function of a guide RNA targeting *Trp53*.

In the second aim, I aim to target our lab's docking mES cell line G4ROSALUC with *in cis* conditional and inducible (COIN) wt/dCas9-MPH containing gene targeting vectors. The COIN wt/dCas9-MPH mES cells when made into a mouse and subsequent breeding of the offspring with tissue specific Cre lines would allow to generate a model expressing Cas9 and MPH in a tissue specific and in an inducible manner. Through this model, simultaneous gene activation and gene knock outs can be achieved.

Chapter 2: Materials and methods

2.1: Mice studies and tumor cell line establishment

All animal work was approved by the University of Manitoba Animal Care Committee. B6;129S4-*Gt(ROSA)26Sor*$^{tm1(rtTA*M2)Jae}$ *Col1a1*$^{tm1(tetO-cas9)Sho}$/J (also known as KH2-iCas9) mice were obtained from The Jackson Laboratory and kept in specific pathogen free (SPF) housing at CancerCare Manitoba. The AAV: ITR-U6-sgRNA(*Kras*)-U6-sgRNA(*Trp53*)-U6-sgRNA(*Lkb1*)-pEFS-Rluc-2A-Cre-shortPA-*KrasG12D*_HDR donor-ITR (also known as AAV-KPL) plasmid (a gift of Feng Zhang, obtained from Addgene) was used by Vector builder to generate adenovirus. Nine female mice were anaesthetized prior to injection of 1 x 10^{11} viral genome copies into the left gastrocnemius muscle via Hamilton syringe. 24 hours later 5 of the mice were provided ad libitum with water supplemented 2 mg/ml doxycycline for 7 days. 2 mice within this cohort were injected intraperitoneally with 1.5 mg/kg doxycycline every other day during this 7-day period.

Tumours were detected by palpation approximately 5 weeks following doxycycline induction and all doxycycline treated mice reached endpoint within 7 weeks of doxycycline induction. None of the mice given straight water reached endpoint during this period.

Upon reaching endpoint, mice were humanely euthanized via CO2 inhalation followed by cervical dislocation.

Visible tumors were excised along with muscle from the non-injected (right) gastrocnemius muscle with no apparent signs of disease (NAD). Livers and spleens were also excised. Each tissue sample (sarcoma and NAD) was divided into 5 parts: one part for DNA extraction, one part for RNA extraction (snap frozen in LN2), one part for protein extraction (snap frozen in LN2), one part for histology (fixed for 24 hours in 4% paraformaldehyde and then kept in 70% EtOH until processing). The remaining sarcoma tissue was ground through a 45 mm cell strainer into complete DMEM to obtain single cell suspensions. After washing the cells with complete DMEM, they were plated into tissue culture treated vessels and allowed to grow until confluent before being frozen in (90% FCS + 10% DMSO) or subjected to protein/RNA/DNA extraction for further analysis.

Fixed tissues were first embedded in paraffin and 5 mm slices placed onto glass slides. H&E staining was performed according to established protocols. The animal work was performed by Dr. Andrew Cuddihy and Katharina Haigh.

2.2: Cell culture: Mouse embryonic stem cell culture

2.21 ES cell culture

For Aim 1, an ES cell line was established from the inducible Cas9 mouse model obtained from Jackson laboratory by our lab manager Katharina Haigh. For Aim 2, F1 G4 hybrid mouse embryonic stem cells (mES) possessing robust developmental potential were utilized for creating genetically engineered mES cells expressing different versions of Cas9 and the MPH gene activation complex.

The ES cells were maintained at 37ºC with 5% humidity. All ES cells were cultured in adherent tissue culture plates that were coated with 0.1 % gelatin and a layer of mitomycin C treated mouse embryonic fibroblasts (MEFs). MEFs were cultured in DMEM (Invitrogen, cat. no. 41965-039), supplemented with 6ml L-glutamine (Invitrogen, cat. no. 25030-024), 6ml 100x pencillin/streptomycin (Invitrogen, cat. no. 15140-122), Beta mercaptoethanol (Sigma, cat. no. M7154), 50ml fetal calf serum (Hyclone, cat. no. CH30160.03), 6ml 100x non essential amino acids (Invitrogen, cat. no. 11140-035) and 6ml 100x sodium pyruvate (Invitrogen, cat. no. 11360-070). mES cells were cultured in knock out DMEM media (Invitrogen, cat. no. 10829-018)) supplemented with 20% FBS, 6ml 100x L-glutamine, 6ml 100x Pencillin/streptomycin, 0.1% 1000x Beta mercaptoethanol, recombinant LIF, 6ml 100x non essential amino acids.

2.22 Passaging of mES cells

mES cells are adherent in nature and were passaged at 1/5, 1/4 or 1/3 dilutions after reaching 70% - 90% confluency. The cells were detached from the plates by incubating with appropriate amount of trypsin for 6-7 minutes at 37ºC. After the brief incubation, the trypsin was neutralized with equal amounts of ES cell media and the cells were seeded into fresh wells/plates for further passaging.

2.23 Cryopreservation of mES cells

Once the mES cells reached 70%-90% confluency, the cells were detached from the plates using trypsin as mentioned above and were pelleted by centrifugation at 1200 rpm for 5 minutes. The cell pellet was washed in cell culture grade PBS and again pelleted by centrifugation at 1200 rpm for 5 minutes. The washed cell pellet usually from a single well of a 6 well plate was resuspended in 3ml of FBS supplemented with 10% dimethyl sulfoxide (DMSO, Sigma Aldritch, USA) which were equally split into 3 cryovials. The cryovials were stored in a Nalgene Mr. Frosty freezing container (Thermofisher scientific USA, which lowers the temperature slowly when frozen at -80ºC for 2-3 days before transferring into liquid nitrogen for long term storage.

2.24 Thawing frozen cells

The frozen cells were taken from liquid nitrogen and thawed in laminar hood, after which they were resuspended in 10ml MEF media to dilute the DMSO and then pelleted by centrifugation at 1200 rpm for 5 minutes. The cell pellet was resuspended in ES cell media and seeded over a confluent layer of Mitomycin C treated feeders.

2.25 ES cell line establishment

For Aim 1, I used an ES cell line that was established in the lab by our lab manager Katharina Haigh from the inducible Cas9 mouse model as according the previously established 2i based protocol[66].

2.3 ES cell DNA isolation

Once the ES cells reach 70%-90% confluency, the cells were pelleted in a 1.5ml Eppendorf and 500 µl of mES cell lysis buffer was added to it along with 2.5µl of proteinase K and was incubated overnight in 55ºC incubator. Then, 200µl of 6M NaCl was added to pellet-lysis buffer mixture and the Eppendorf was constantly shook for 5 minutes followed by centrifugation at maximum speed for 10 minutes. The supernatant was collected and equal amount of isopropanol was added followed by constant shaking for 2-5 minutes till shiny thread like structures are seen and centrifuged at maximum speed for 5 minutes. The supernatant was discarded and the DNA pellet was washed in 70% ethanol and then air dried till ethanol residues was completely removed. The DNA pellet was then resolved in 200µl of nuclease free water and the concentration was calculated using nanodrop. The resolved DNA was then stored either at 4ºC or at -20ºC for long term storage.

2.4 *Trp53* targeting gRNA containing plasmid

A plasmid containing a guide targeting *Trp53* was obtained from Sigma. The plasmid consists of Rous sarcoma virus long terminal repeat (RSV LTR) followed by a rev response element that has a crucial role in packaging the virion particles[67]. This sequence is followed by a U6 promoter that drives the expression of a 20 bp guide RNA targeting mouse *Trp53* gene followed by a central poly purine tract/central termination sequence (cPPT) that helps to increase the integrating and transducing capacity of the viral vector[68]. cPPT sequence is followed by human elongation factor 1 alpha (Ef1a) promoter driving the expression of a puromycin resistance gene and blue fluorescent protein (BFP) separated by a P2A sequence. P2A peptide site allows two separate proteins to be generated from the fused protein by endogenous proteases[69].This element is followed by a WPRE and a 3' LTR sequence. Woodchuck hepatitis virus post transcriptional regulatory element (WPRE) enhances the expression of genes present in a viral vector[70].

2.5 Lentivirus Production

Human embryonic kidney 293 cells containing the SV40 Large T antigen (HEK293T) were maintained in DMEM supplemented with 10% heat inactivated FCS and Pen/Strep. 24 hours prior to transfection, 4 million cells were plated in a 10 cm dish. The day of transfection, 5 mg of either Cppt2E (gift of Sam Kung, University of Manitoba) or the *Trp53* sgRNA vector was combined with 2 mg of the psPAX2 lentiviral packaging plasmid (a gift from Didier Trono) and 1 mg of the amphitropic VSV-G envelope expressing plasmid PMD2.G (a gift from Didier Trono) in 1 ml of serum free DMEM. 24 ml of 1 mg/ml polyethyleneimine (PEI, Polysciences) was added to the diluted DNA, vortexed and incubated at room temperature for 15 minutes before being added dropwise to cells. 24 hours later, transfection media was removed and replaced with fresh complete DMEM. Lentiviral supernatant was collected after an additional 24 hours, syringe filtered thought a 45-micron filter before being added to mESCs. Lentiviral packaging was performed by Dr. Aissa Benyoucef

2.6 Quantitative Reverse transcription polymerase chain reaction

The total RNA was isolated according to the instructions provided in the RNA isolation kit (GE Health Care, RNA spin Mini, Cat no: 45001162). cDNA was isolated using first strand cDNA synthesis kit – Transcript II – All in one first strand cDNA synthesis super mix for qPCR from Trans gen biotech (cat no: AT341) by using 1µg of total mRNA. qRT-PCR was performed using the perfect start green qPCR Super mix in a light cycler. Gene expression was standardized against beta-actin. All the reactions were run in technical triplicates and biological replicates.

2.7 Multi site Gateway cloning - (L-R reaction)

Multisite Gateway cloning based on L-R reaction was the methodology that was employed to create the gene targeting vector for genetically manipulating mES cells. Three different entry vectors and one destination vector were used to create the wt/dCas9-MPH gene targeting vector. All the entry vectors and the destination vector are flanked with Gateway cloning compatible L-R recombination sites which facilitates recombination between similar L-R sites containing plasmids.

100ng of each cDNA containing entry vectors were utilized for one reaction. In the first step of the Gateway cloning, 100ng of each of the three entry vectors were mixed together in a single reaction along with 2µl of LR clonase enzyme. The total reaction volume for this step was made to 8µl by adding nuclease free water and then the reaction mixture was incubated at 25ºC for 8 hours. After 8 hours of incubation, 100 ng of the destination vector is added to the same mixture along with 2 µl of clonase and the final volume was made to 15 µl with nuclease free water. This reaction mixture was

incubated overnight at 25ºC to generate the different versions of wt/Cas9-MPH gene targeting vector. The gene targeting vector was transformed into competent cells to generate the expression vector for targeting into mES cells.

2.8 Targeting different wt/dCas9-MPH vectors into mES cells

Different variants of the wt/dCas9-MPH gene targeting vector that were generated and validated were targeted into the docking ES cells using lipofectamine transfection. The targeting conditions for one well of a 6 well plate containing the docking mES cells was performed by co transfecting the targeting vector with a FlpE expressing plasmid (pCAG-FlpE) in a 1:1 ratio along with 7ul of Lipofectamine (Lipofectamine™ 3000 Transfection Reagent, Thermofisher Cat no L3000001). After 24 hrs of transfection, the mES cells from a single well of a 6 well plate was trypsinized and transferred to a new 10 cm MEF coated tissue culture dish and G418 drug selection (225µg/ml) was started the following day. The drug selection is usually followed for a period of 10-12 days after which the individual surviving colonies are picked manually by using a 2-20µl pipette under a microscope. The individual picked colonies were transferred into single wells of a 96 well plate containing 40 µl of 1X PBS and then trypsinzed into 24 well plates for expansion. Colonies from 24 well plate were expanded and DNA was isolated from them for PCR for the correct integration suing the R25 5' Integration PCR. Forward primer (5' to 3') AAAGTCGCTC TGAGTTGTTAT Reverse primer (5' to 3') GCGGCCTCGACTCTACGATA

2.9 Competent cell transformation

The required number of competent cell vials were transferred from the -80ºC and thawed on ice. 1µl of the reaction mixture (usually multi site Gateway cloning L-R reaction mixture) was gently added to the competent cell mixture and incubated on ice for 30 minutes. The competent cell mixture was then given rapid heat shock at 42ºC for 45 seconds followed by incubation on ice for 1 minute. Then 500 µl of SOC media was added to the tubes and incubated for 1 hour at 37ºC in a constant shaker at 200 rpm. Then, 50-100 µl of transformed cell suspension was streaked over appropriate agar plates containing the antibiotic selection marker and sealed with a paraffin strip and incubated at 37ºC incubator overnight. The transformed colonies were individually picked and grown in 2ml LB media with the antibiotic resistance (Mini Prep).

2.10 Plasmid DNA isolation

The plasmid DNA from the Mini Prep was isolated using the Promega kit method and were quantified using Nanodrop. In order to scale up the plasmid, Midi Prep was made by adding some liquid from the

LB broth used for Mini prep into 100 ml of LB broth with appropriate antibiotic resistance and incubated at 37ºC incubator overnight with a constant shaking. Similarly, the plasmid DNA from Midi prep was extracted using the Promega kit-based method and quantified using Nanodrop. The plasmid DNA was stored at -20ºC for long term storage.

2.11 Restriction digestion

The transformed colonies were restriction digested for correct size and orientation. Restriction digestion was usually carried out by digesting 1µg of the plasmid in 1µl of the desired restriction enzyme or as directed in the appropriate protocol with the optimal type of the buffer. The digested samples were validated by running them on an agarose gel for 30-40 minutes in 0.5 X TAE buffer.

2.12 PCR product isolation

The PCR product was isolated by using a Promega Wizard SV gel and PCR clean up system. Equal amount of membrane binding solution was added to the PCR amplification and the prepared product was transferred to the SV Mini column assembly and incubated for 1 minute at room temperature. The SV Mini column assembly was centrifuged for 1 minute at 16,000g and the collection tube was emptied. The SV Mini column was returned to the collection tube and washed by adding 700ul of 95% ethanol and then centrifuged for 1minute at 16,000 g. The collection tube was emptied and the column was again washed by adding 500ul of 95% ethanol and the SV column assembly was centrifuged for 16,000 g for 5 minutes. The collection tube was emptied and then the SV column assembly was centrifuged for 1 minute at 16,000 g to remove any residual ethanol. The collection tube was then discarded and the Mini column was placed into a fresh Eppendorf and 20-28 ul of nuclease free water was added to the SV Mini column and incubated for 1 minute at room temperature. Then the tubes were then centrifuged at 16,000 g for 1 minute to elute the purified PCR product into the Eppendorf which was stored at either 4ºC for short term or at -20ºC for long term storage.

2.13 DNA Sequencing

The restriction digestion validated wt/dCas9-MPH gene targeting vectors generated through L-R cloning-based Gateway cloning setup were sequence verified through Sanger sequencing. It was done by splitting the entire wt/dCas9-MPH gene targeting vector into 21 different fragments and designing sequencing primers that amplify around 800 bp of each of the 21 regions with an overlap of 100 bp between two consecutive regions. Using these primers, the entire gene targeting vector was PCR amplified, purified through the above-mentioned protocol and then sent for Sanger sequencing. The putative gene targeting vector sequence was obtained by performing a simulation of the multisite

Gateway cloning of the three different entry vectors with the destination vector using the SnapGene software which resulted in the generation of the targeting vector. The sequence obtained from the sanger sequencing were retrieved using the Finch software and they were aligned with the targeting vector sequence from the Snapgene using the NCBI blastn software to analyze any mutations.

SN	Location (Conditional and inducible wt/dCas9-MPH sequence)	Left Primer	Right Primer	Regions covered	Product Size
1.	1 to 1000 bp	tcgcgactcgaggaataaat	ctacccgcttccattgctc	Insulator, PGK promoter	827
2.	900 to 1900 bp	gccaccatggagaagttcct	agtcgtgtcttaccgggttg	PGK promoter, ORI	840
3.	1800 to 2800 bp	gcagccactggtaacaggat	gaggaccgaaggagctaacc	ORI, AmpR	850
4.	2700 to 3700 bp	aagtaagttggccgcagtgt	gcctgatgcggtattttctc	AmpR, Frt sites	845
5.	3600 to 4600 bp	gttttcccagtcacgacgtt	acagaggctgctgatctcgt	Frt wt and loxp sites	830
6.	4500 to 5500 bp	ttgccaagttctaattccatca	attaagggttccggatcagc	Region between two loxp sites	843
7.	5400 to 6400 bp	gagctcgttttcccagtcac	cctcgatggtagacccgtaa	One loxp sites and rtTA	835
7.b	5300 to 6500 bp	agctgatccggaacccttaa	ctctgcaccttggtgatcaa	One loxp sites and rtTA	
8.	6300 to 7300 bp	gcgcgtacgaaaaacaattac	ccccagatcagatcccatac	Part rtTA, full IRES, starting of puro	816
8b.	6250 to 7250 bp	gcgcgtacgaaaaacaattac	ccccagatcagatcccatac	Part rtTA, full IRES, starting of puro	816
9.	7200 to 8200 bp	ggacgtggttttcctttgaa	actcagacaatgcgatgcaa	Full Puro and Insulator	846
9b.	7100 to 8300 bp	tatgggatctgatctggggc	gcaactagaaggcacagtcg	Full Puro and Insulator	825

9c.	7150 to 8150 bp	ggacgtggttttcctttgaa	actcagacaatgcgatgcaa	Full Puro and Insulator	846
10.	8100 to 9100 bp	cttctgaggcggaaagaacc	ggtaccgagctcgactttca	Insulator, TRE sequence	890
11.	9000 to 10,000 bp	catccacgctgttttgacct	tgtgtagacttgggcactgg	Part TRE, MS2 and part p65	834
12.	9900 to 10,900 bp	gctccagtgcccaagtctac	aggattctcctcgacgtcac	Full p65, HSF1 and part Cas9	846
13.	10,800 to 11,800 bp	caaagttctgggcaataccg	gctttggtgatctccgtgtt	Cas9	849
14.	11,700 to 12,700 bp	aacacggagatcaccaaagc	gttgaagcgatcctccactc	Cas9	840
15.	12,600 to 13,600 bp	ggacttcctggacaatgagg	tgccgccaataatttttcat	Cas9	843
15.b	12,550 to 13,550 bp	caaggacttcctggacaatga	ccgttgtgtgatcagtttgg	Cas9	877
16.	13,500 to 14,500 bp	ttcacgcatgaacaccaagt	gcactagcgagcattcgttt	Cas9	843
17.	14,400 to 15.400 bp	ggaaacgaatgctcgctagt	aggaactgcttccttcacga	Part Cas9 and IRES	809
17.b	14,300 to 15,500 bp	agcttcccaagtactctctct	ttgcattcctttggcgagag	Part Cas9 and IRES	815
17.c	14,350 to 15,350 bp	ggaaacgaatgctcgctagt	aggaactgcttccttcacga	Part Cas9 and IRES	809
18.	15,300 to 16,300 bp	tggctctcctcaagcgtatt	ctgggtgctcaggtagtggt	Part IRES, full EGFP	808
19.	16,200 to 17,200 bp	gtcctgctggagttcgtgac	aaaaccgtgatggaatggaa	Part EGFP, Luciferase	822
20.	17,100 to 18,100 bp	aagattcaaagtgcgctgct	ggccgctctagaattacacg	Part Luciferase and part PA	826
21.	17,900 to 18,840 bp	ggaaaactcgacgcaagaaa	aggaaaggacagtgggagtg	PA and insulator sequence	848

21.b	17,850 to 18,840 bp	gtggacgaagtaccgaaagg	gcgatgcaatttcctcattt	PA and insulator sequence	899

2.14 Flow cytometry analysis

mES cells that are correctly targeted with the appropriate gene targeting vector would be expressing enhanced green fluorescent protein (EGFP) from the R26 locus and the levels of EGFP were quantified through flow cytometry analysis using the flow cytometer (Novocyte, Novosampler Pro). The desired mES cells from the tissue culture plates were pelleted and they were washed twice with 1x PBS and were resuspended in fresh 560 µl of 1x PBS and they were analyzed for EGFP expression by running the samples on FITC channel in the No The gates were set by eliminating debris as well as using our lab' s docking mES cell line G4ROSALUC that does not contain the EGFP cassette in it as the negative control.

2.15 X-gal staining of cells

mES cells were fixed by adding 0.2 % of Glutaraldehyde for 10 minutes and washed for 10 minutes in X gal wash buffer (100mM sodium phosphate, 2mM MgCl2, 0.01 % sodium deoxycholate, 0.02% Np40). Then, 2 ml of X gal staining solution (0.5 mg/ml X-gal, 5 mM potassium ferrocyanide and 5mM potassium ferricyanide) in X gal wash buffer was added to the wells and incubated for 1 hr at 37ºC incubator protected from light. The stained cells could be seen under the naked eye and the cells could be imaged under a bright field microscope.

2.16 T7 Endonuclease assay

Genomic editing created by the guide RNA was validated for cleavage using the T7 endonuclease assay. It is an assay that helps to detect the Cas9 mediated edits by cleaving the mismatches that usually occurs after Cas9 creates double stranded breaks. The assay involves designing primers that amplifies (around 800 bp) the region surrounding the putative target site of the guide RNAs. PCR amplification is done using the designed primers to amplify the targeted region in the desired cells. The PCR product is purified and 2µg of the purified product is added to a reaction mixture consisting of 2µl NEB buffer 2 and nuclease free water made up to 19µl. The reaction mixture is denatured at 95ºC for 5 minutes in a thermal amplifier followed by cooling in the same cycler just by opening the lids for 20 minutes. To the cooled reaction mixture, 1µl of T7 endonuclease enzyme (NEB, Catalogue # M0302S) is added and incubated at 37ºC for 30 minutes in a thermal cycler. After 30 minutes, the product is run on an 1.5% agarose gel and the gel is imaged to identify the cleaved bands.

Chapter 3: Results

Part 1: 3.1 To utilize a doxycycline-inducible Cas9 mouse model and ES cell derivatives for sarcoma modeling and *Trp53* targeting gRNA validation respectively

3.1.1 Background

In 2020, Bowling et al., developed a mouse model which expresses Cas9 in a doxycycline (DOX) dependent manner[71]. This mouse model expresses rtTA from the *R26* locus and wtCas9 from the *Col1a1* locus placed under the transcriptional control of the Tet- responsive element (TRE promoter: Tet (o)-CMV_{min}). When DOX is administered to this mouse via their drinking water, rtTA binds with DOX which in turn leads to rtTA translocation to the nucleus and binds to the TRE promoter to activate the expression of Cas9. This mouse model allows for temporal control of Cas9 expression thereby reducing potential off-target effects associated with constitutive expression of Cas9[72]. By breeding this model, with another mouse model expressing sgRNA cassette with a unique zebrafish gRNA target array from the *Col1A1* locus, the same group developed a double-transgenic mouse model that was combined with single-cell RNAseq approaches to simultaneously understand the lineage history and the gene expression changes of single cells over a period of time by tracking the mutations introduced by the sgRNA and Cas9[71] in all subsequent progeny cells following pulsed DOX-mediated expression of Cas9. This approach illustrated a novel methodology for defining the *in vivo* identity of different cell fate at the single cell level even in the absence of cells surface marker expression as they are identified based on their gene expression changes. Application of this model elucidated large differences in hematopoietic stem cell (HSC) clone distribution across different bones marrow niches. Furthermore, this study also revealed that the capabilities of HSCs to give rise to different clones dampens greatly after treatment with 5' FC. This mouse also shed light on a key gene that acts as the switch between quiescent and proliferating HSCs[71].

We obtained the ***Gt (ROSA)26Sor$^{tm1(rtTA*M2)Jae}$ Col1a1$^{tm1(tetO-cas9)Sho}$*** **mice (iCas9 mice)** from the Jackson Laboratory and established embryonic stem (ES) cell lines from the homozygous blastocysts[71] in order to validate DOX-mediated Cas9 expression and to utilize these ES cells as an *in vitro* platform for validating the functionality of different guide RNAs before using them for *in vivo* applications.

3.1.2 Establishment and validation of mES cells from the iCas9 mouse

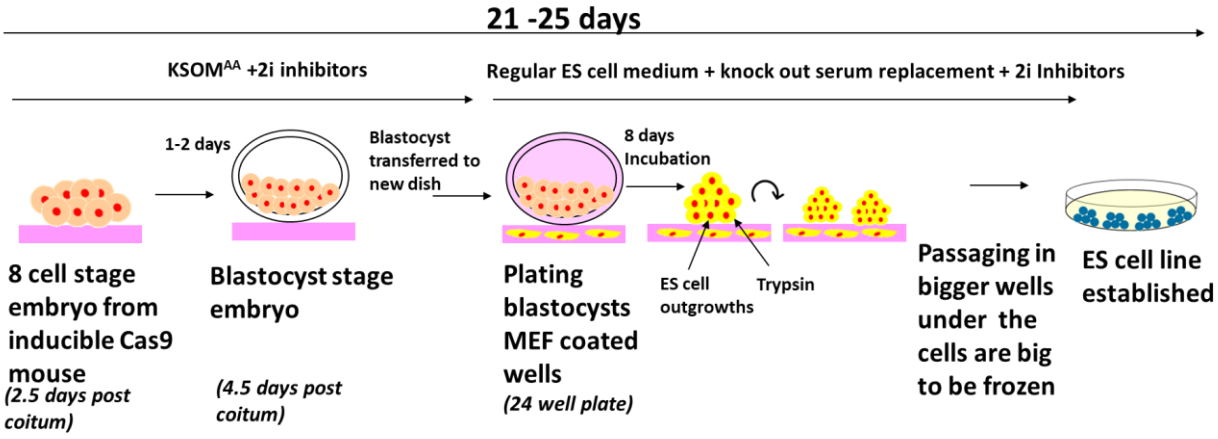
The first aim was to establish an *in vitro* platform using mES cells derived from an iCas9 mouse model wherein the efficacy of different guide RNAs could be tested *in vitro* before utilizing them for *in vivo* studies. mES cells were established from the iCas9 mouse obtained from Jackson's laboratory (with

the assistance of Katharina Haigh, Lab manager of the Haigh lab) and are denoted as iCas9 mES cells in further sections of Aim 1 **(Figure 3.1 A).** Twelve iCas9 mES cell lines were established using the KSOM + two inhibitors (2i) based protocol. One inhibitor is PD0325901 which is a MAPK or ERK inhibitor and suppresses the differentiation inducing signaling. The other inhibitor is CHIR99021 which is a GSK-3 inhibitor and promotes the proliferation and survival of mES cells[66]. Since, the protocol entails the use of both differentiation inhibiting and proliferating promoting signals, the success of establishing mES cells becomes quite high[66] (**Figure 3.1 B)**. An ES cell line that was growing well was chosen and used for subsequent studies. **Figure 3.1 C** depicts an established iCas9 ES cell line with round shaped colony morphology plated on a layer of mouse embryonic fibroblasts. The iCas9 mES cell system has the Cas9 gene placed under the transcriptional control of the TRE promoter expressed from the *Col1A1* locus and the reverse tetracycline transactivator (rtTA) expressed from the endogenous *R26* locus **(Figure 3.2 A).** A passage-1 (P1) ES cell line was cultured and two different passages (P2 and P3) were sub cultured from it and P1, P2 and P3 were used for detecting Cas9 expression upon DOX induction for 24, 48, 72 and 96 hrs. mRNA was isolated from both the control and DOX treated ES cells groups and the expression levels of Cas9 were evaluated through qPCR with non-DOX treated ES cells as control and normalized to beta actin expression levels. mRNA quantification revealed a three-hundred-fold increase in Cas9 expression in the iCas9 ES cells after 24 hrs of doxycycline induction compared to the non-DOX induced iCas9 ES cells **(Figure 3.2 B)**. The highest expression of DOX mediated Cas9 induction was observed 24 hrs after DOX induction and the expression levels of Cas9 induction though remained high but steadily decreased after 24 hrs of DOX induction probably due to the epigenetic silencing of the minimal CMV promoter that is contained within the TRE[26,27].

A.

Gt (ROSA)26Sor$^{tm1(rtTA*M2)Jae}$ *Col1a1*$^{tm1(tetO-cas9)Sho}$ (Inducible Cas9 mouse)

B.



C. **iCas9mES cell line # 1**

Bright field microscope Image

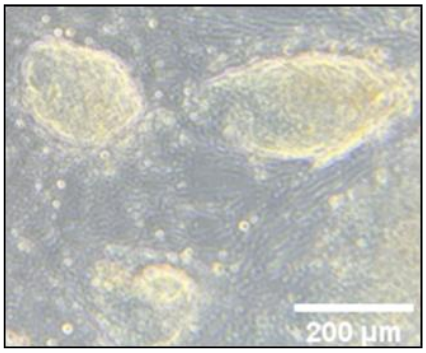
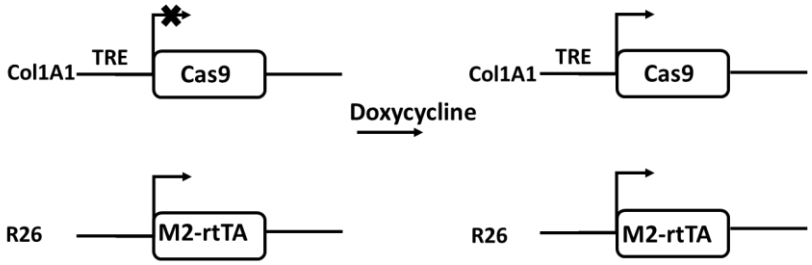


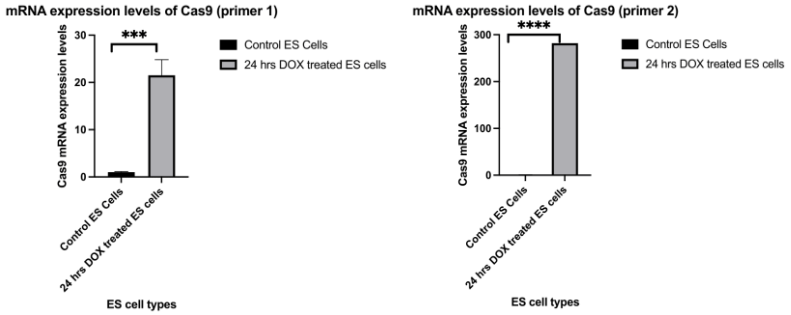
Figure 3.1: Establishment of ES cell lines from *Gt (ROSA)26Sor*$^{tm1(rtTA*M2)Jae}$ *Col1a1*$^{tm1(tetO-cas9)Sho}$ mice (iCas9 mouse) model using the 2i protocol

A. Breeding scheme depicting the generation of the ***Gt (ROSA)26Sor***$^{tm1(rtTA*M2)Jae}$ ***Col1a1***$^{tm1(tetO-cas9)Sho}$ **mice (iCas9 mouse)**. The ***Gt (ROSA)26Sor***$^{tm1(rtTA*M2)Jae}$ mouse expressing constitutive rtTA from the R26 locus was previously bred with the ***Col1a1***$^{tm1(tetO-cas9)Sho}$ mouse wherein Cas9 expression is under the control of TRE promoter (tet (o)-CMV_{min}) to generate the ***Gt (ROSA)26Sor***$^{tm1(rtTA*M2)Jae}$ ***Col1a1***$^{tm1(tetO-cas9)Sho}$ **mice (iCas9 mouse)** double homozygous mouse. The **iCas9 mouse** line is maintained in double homozygous breeding pairs and is expected to express Cas9 from the *Col1a1* locus when it is administered doxycycline. **B.** ES cells were established from the iCas9 blastocysts using the two inhibitors: ERK inhibitor (differentiation inhibitor) and GSK-3 inhibitor (proliferation promoter) (2i) based protocol. Morula stage embryo (8 cell stage embryo) were allowed to develop into blastocysts in a KSOM + 2i media. The blastocysts were then plated onto murine embryonic fibroblast (MEF) coated dishes and the blastocyst outgrowths were disrupted with trypsin and plated into fresh MEFs to develop an ES cell line. **C.** Bright field white light images of an iCas9 ES cell line. The iCas9 mES cells were grown on a layer of MEFs. The scale bar for the brightfield image is 200 μM.

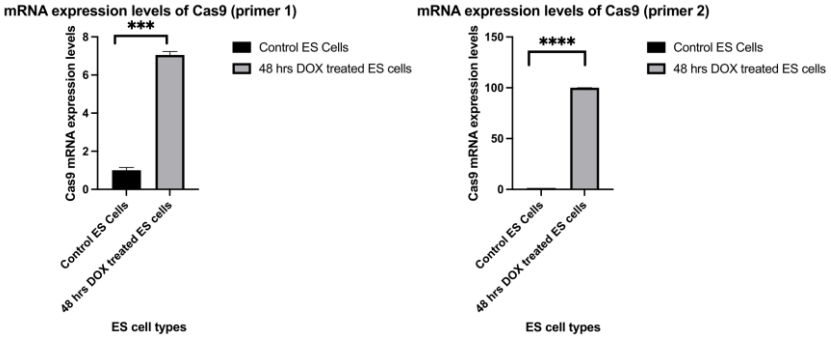
A.


Col1A1
TRE
Cas9
Doxycycline
Col1A1
TRE
Cas9
R26
M2-rtTA
R26
M2-rtTA

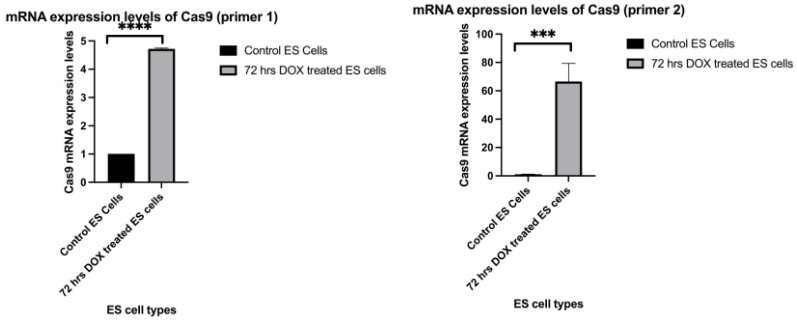
B. i.


mRNA expression levels of Cas9 (primer 1)
***
Control ES Cells
24 hrs DOX treated ES cells
Cas9 mRNA expression levels
Control ES Cells
24 hrs DOX treated ES cells
ES cell types
mRNA expression levels of Cas9 (primer 2)
****
Control ES Cells
24 hrs DOX treated ES cells
Cas9 mRNA expression levels
Control ES Cells
24 hrs DOX treated ES cells
ES cell types

ii.


mRNA expression levels of Cas9 (primer 1)
***
Control ES Cells
48 hrs DOX treated ES cells
Cas9 mRNA expression levels
Control ES Cells
48 hrs DOX treated ES cells
ES cell types
mRNA expression levels of Cas9 (primer 2)
****
Control ES Cells
48 hrs DOX treated ES cells
Cas9 mRNA expression levels
Control ES Cells
48 hrs DOX treated ES cells
ES cell types

iii.


mRNA expression levels of Cas9 (primer 1)
****
Control ES Cells
72 hrs DOX treated ES cells
Cas9 mRNA expression levels
Control ES Cells
72 hrs DOX treated ES cells
ES cell types
mRNA expression levels of Cas9 (primer 2)
***
Control ES Cells
72 hrs DOX treated ES cells
Cas9 mRNA expression levels
Control ES Cells
72 hrs DOX treated ES cells
ES cell types

iv.

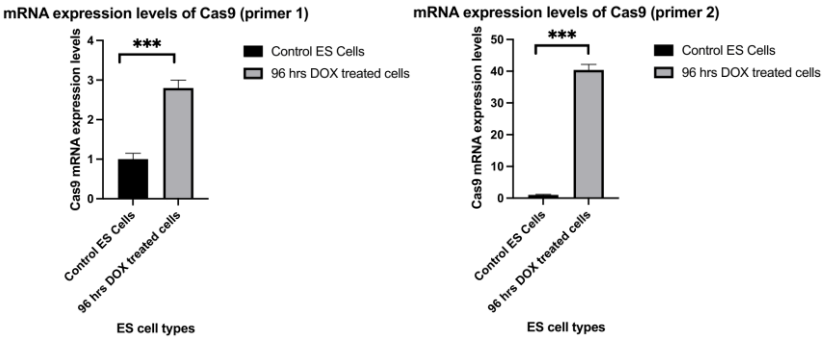

mRNA expression levels of Cas9 (primer 1)
***
Control ES Cells
96 hrs DOX treated cells
Cas9 mRNA expression levels
Control ES Cells
96 hrs DOX treated cells
ES cell types
mRNA expression levels of Cas9 (primer 2)
***
Control ES Cells
96 hrs DOX treated cells
Cas9 mRNA expression levels
Control ES Cells
96 hrs DOX treated cells
ES cell types

Figure 3.2: qPCR analysis illustrates Cas9 expression induced in the iCas9 mES cells only when treated with doxycycline

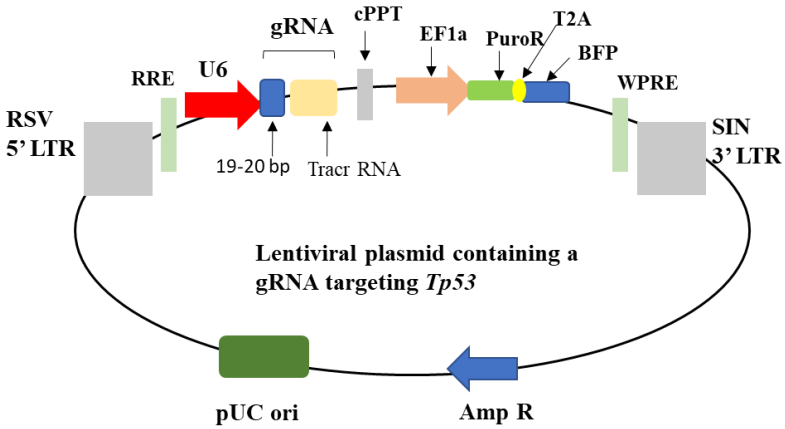
A. The iCas9 inducible system consists of the Cas9 whose expression is under the control of the TRE promoter (TRE: (tet (o) – CMV_{min}) in the Col1A1 locus. The revere tetracycline controlled transactivator (rtTA) expression is driven by the ROSA 26 promoter. When doxycycline (DOX) is administered to the system, rtTA forms a complex with DOX and translocate to the nucleus and binds to the TRE promoter driving the Cas9 expression. The expression of Cas9 is thus dependent upon the administration of doxycycline. **B.** Quantification of Cas9 mRNA expression levels in an iCas9 ES cell line using two different Cas9 primers under DOX induced and non-induced conditions revealed the robust expression levels of Cas9 after 24 hrs of DOX induction (B: i) which steadily decreased until the last experimental time period (B: ii:48 hrs, iii: 72 hrs, iv: 96 hrs). Number of samples (N) = 3 (three different passages), number of replicates per timepoint from each passage (n) = 3, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p<0.001$. Statistical significance was performed by using student t test with two tailed distributions.

3.1.3 iCas9 mES cell system as an *in vitro* platform to validate the function of a commercial guide RNA targeting *Trp53*

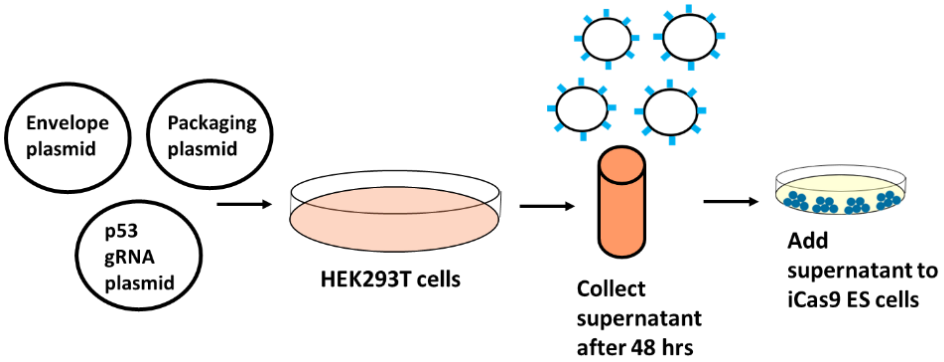
Transduction of a plasmid containing guides targeting *Trp53* **(Figure 3.3 A)** packaged into a lentivirus into **(Figure 3.3 B)** iCas9 mES cells and after 48 hrs, the cells were split into 2 different groups. 2µg/ml doxycycline was added to the one batch of cells for 4 days and fluorescence imaging revealed a strong BFP expression **(Figure 3.3 B)** only in the cells transduced with the virus. Also, another batch of iCas9 ES cells that was not transduced with the virus but treated with DOX media was used as the real negative control. Fluorescence microscopy images were taken for all the three batches of iCas9 ES cells once they were confluent enough to observe the expression of BFP. Through fluorescence microscopy images, it was observed that the BFP expression was strongly observed only in the cells that were transduced with the virus

DNA was isolated from all the 3 groups of iCas9 ES cells (iCas9 ES cells plus DOX, iCas9 ES cells transduced with lentivirus containing guides targeting *Trp53*, iCas9 ES cells transduced with lentivirus containing guides targeting *Trp53* and also induced with DOX for 96hrs) **(Figure 3.4 A).** T7 Endonuclease assay that determines the mismatches created by Cas9-gRNA complex was used to validate the function of the induced Cas9. First, a pair of primers that specifically amplifies around 800bp surrounding the guide RNA target site in the mouse *Trp53* gene was designed through primer 3 software. **(Figure 3.4 B)** PCR was performed on genomic DNA obtained from all the three groups of iCas9 ES cells (iCas9 ES cells treated with DOX only, iCas9 ES cells treated with the viral supernatant alone, iCas9 ES cells treated with viral supernatant and also treated with DOX). The PCR amplified products were purified, denatured, reannealed, subjected to the action of the T7 endonuclease and the digested products were run in an 1.5% agarose gel. Gel electrophoresis **(Figure 3.4 C)** revealed that the DNA-cleavage is observed in the PCR products derived from the iCas9 ES cells that was transduced with the virus and also induced with DOX continuously for four days. The Cas9-mediated DNA cleavage was demonstrated by the presence of two extra bands that are less than 800bp (cleaved bands) in addition to the main band observed at 800bp. Evidence of mild DNA-cleavage was also observed in the cells that were transduced with virus and not induced with DOX. This might have occurred because of a mild basal level of Cas9 mRNA expression observed even under non-induced state.

A.


gRNA
cPPT
EF1a
PuroR
T2A
BFP
RRE
U6
WPRE
RSV
5' LTR
SIN
3' LTR
19-20 bp
Tracr RNA
Lentiviral plasmid containing a
gRNA targeting *Tp53*
pUC ori
Amp R

B.


Envelope plasmid
Packaging plasmid
p53 gRNA plasmid
HEK293T cells
Collect supernatant after 48 hrs
Add supernatant to iCas9 ES cells

C.

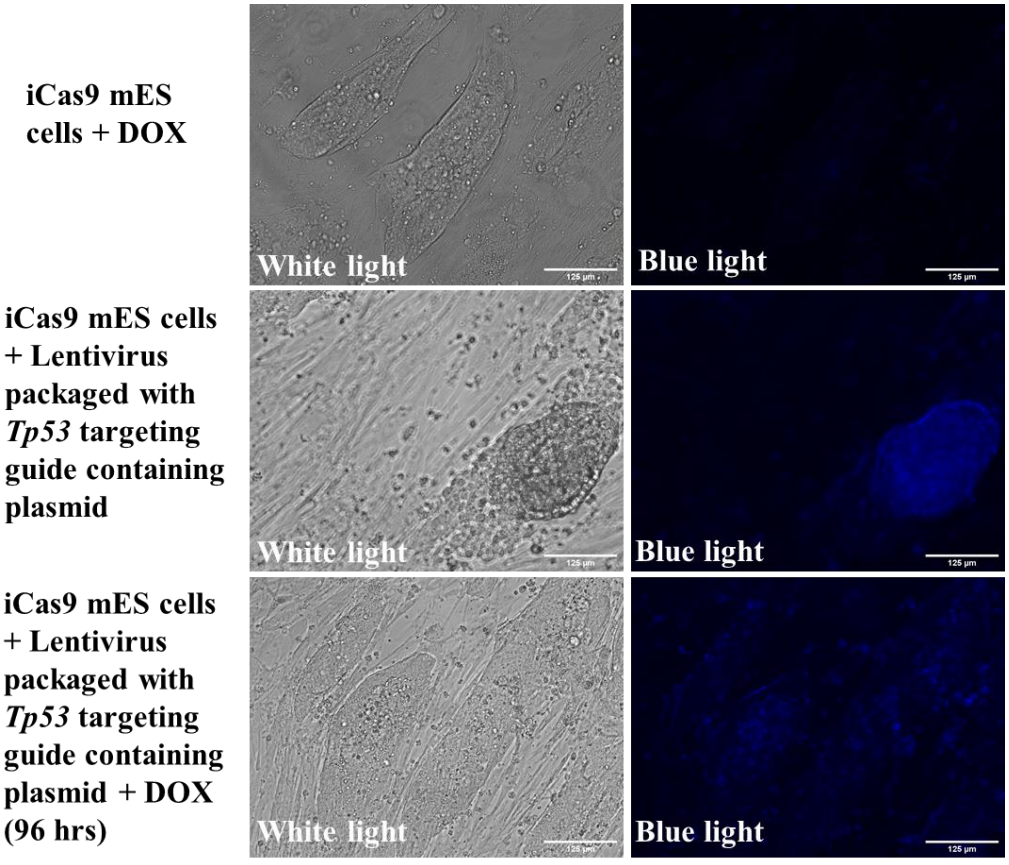
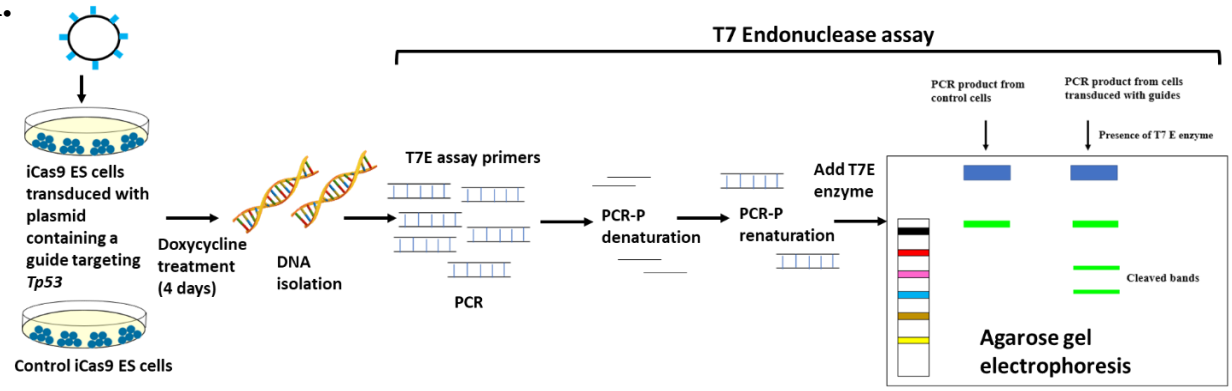

iCas9 mES cells + DOX
White light
Blue light
iCas9 mES cells + Lentivirus packaged with *Tp53* targeting guide containing plasmid
White light
Blue light
iCas9 mES cells + Lentivirus packaged with *Tp53* targeting guide containing plasmid + DOX (96 hrs)
White light
Blue light

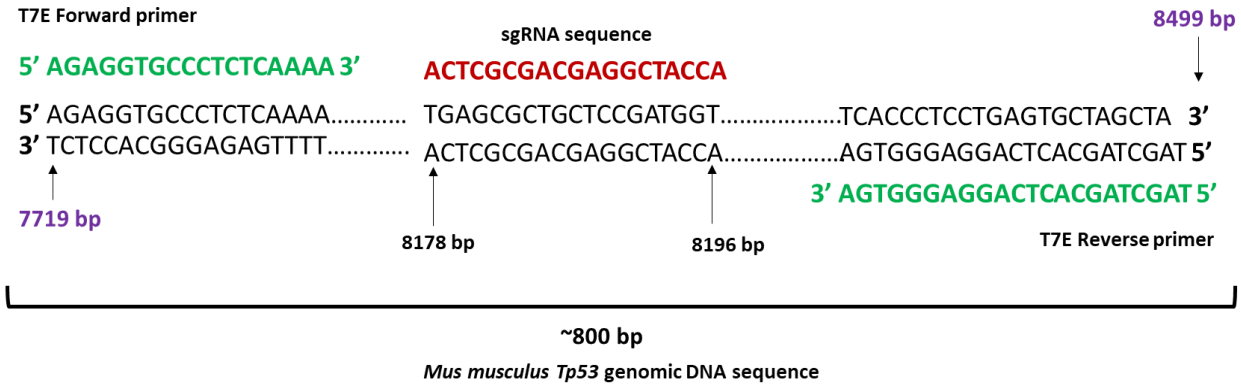
Figure 3.3: Lentiviral transduction of the plasmid containing a gRNA targeting *Trp53* into iCas9 mES cells

A. Schematic of the lentiviral plasmid containing the guide targeting *Trp53*. The plasmid consists of Rous sarcoma virus long terminal repeat (RSV LTR) followed by a rev response element that has a crucial role in packaging the virion particles. This is followed by a U6 promoter that drives the expression of a 20bp guide RNA targeting mouse *Trp53* followed by a central poly purine tract/central termination sequence (cPPT). cPPT sequence is followed by a human elongation factor 1 alpha (Ef1a) promoter driving the expression of a puromycin resistance gene and blue fluorescent protein separated by a P2A sequence. This element is followed by a Woodchuck hepatitis virus post transcriptional regulatory element (WPRE) and 3' LTR sequence. **B.** Schematic workflow depicting the process of packaging the plasmid into lentivirus for transduction into iCas9 mES cells. Human embryonic kidney cells (HEK 293) were cultured to near confluency. The guide RNA containing plasmid, the envelope expressing plasmid and the packaging plasmids were mixed together and added to the HEK cells. The viral supernatant was collected at desired time periods and then syringe filtered and transduced into iCas9 ES cells. **C.** Fluorescence microscopy images of different iCas9 mES cell groups. Strong BFP expression is only observed in the iCas9 ES cells that was transduced with the virus and/ induced with doxycycline. The scale bar represents 125 μm.

A.


T7 Endonuclease assay
iCas9 ES cells transduced with plasmid containing a guide targeting *Tp53*
Control iCas9 ES cells
Doxycycline treatment (4 days)
DNA isolation
T7E assay primers
PCR
PCR-P denaturation
PCR-P renaturation
Add T7E enzyme
PCR product from control cells
PCR product from cells transduced with guides
Presence of T7 E enzyme
Cleaved bands
Agarose gel electrophoresis

B.


T7E Forward primer
5' AGAGGTGCCCTCTCAAAA 3'
sgRNA sequence
ACTCGCGACGAGGCTACCA
8499 bp
5' AGAGGTGCCCTCTCAAAA..... TGAGCGCTGCTCCGATGGT.....TCACCCTCCTGAGTGCTAGCTA 3'
3' TCTCCACGGGAGAGTTTT..... ACTCGCGACGAGGCTACCA.....AGTGGGAGGACTCACGATCGAT 5'
3' AGTGGGAGGACTCACGATCGAT 5'
7719 bp
8178 bp
8196 bp
T7E Reverse primer
~800 bp
Mus musculus Tp53 genomic DNA sequence

C.

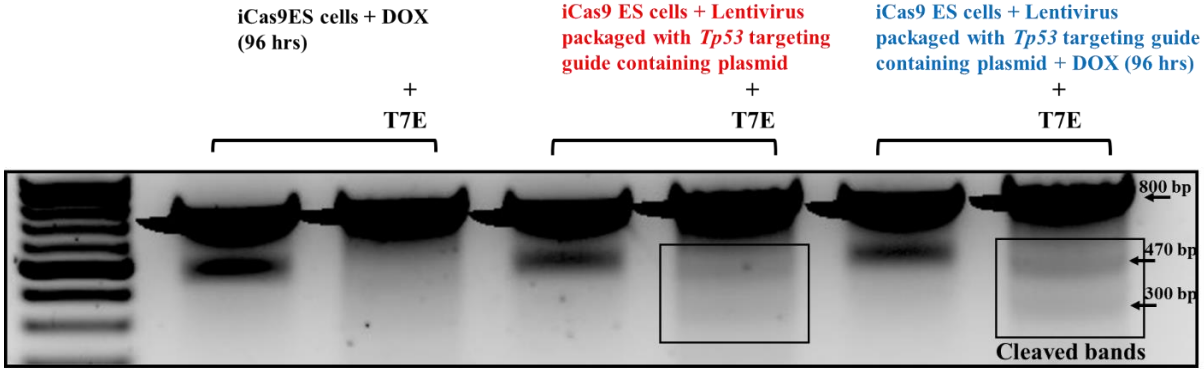

iCas9ES cells + DOX (96 hrs)
+
T7E
iCas9 ES cells + Lentivirus packaged with *Tp53* targeting guide containing plasmid
+
T7E
iCas9 ES cells + Lentivirus packaged with *Tp53* targeting guide containing plasmid + DOX (96 hrs)
+
T7E
800 bp
470 bp
300 bp
Cleaved bands

Figure 3.4: T7 endonuclease assay reveals the mutations created by Cas9-gRNA in the *Trp53* gene of the iCas9 mES cells transduced with a guide targeting *Trp53* and induced with DOX

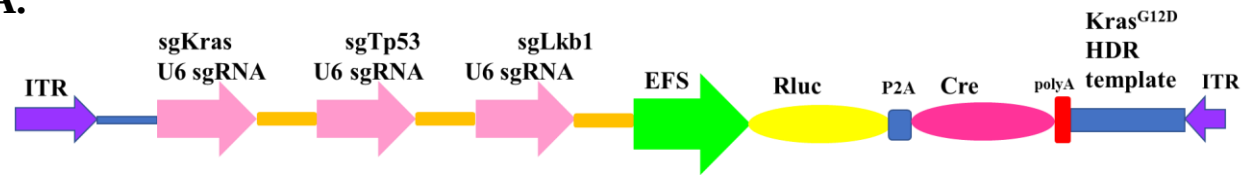
A. The schematic of the T7 Endonuclease assay. DNA is extracted from different iCas9 mES cell groups (iCas9 ES cells treated with DOX only, iCas9 ES cells transduced with the viral supernatant alone, iCas9 ES cells transduced with viral supernatant and also treated with DOX for 96 hrs) were subjected to a PCR by designing a pair of primers that specifically amplifies 800 bps surrounding the guide RNA target region in the mouse *Trp53* gene. The PCR amplified products were purified, denatured, reannealed, subjected to the action of the T7 endonuclease enzyme and the digested products were run in an 1.5% agarose gel. The T7 endonuclease enzyme cleaves any mismatches and the presence of cleaved/ extra bands would be an indicator of the genetic alterations created by the Cas9-gRNA complex. **B.** The sgRNA target site in the *Trp53* gene and the primers used for amplifying 800 bp surrounding the target region Forward (Fw) primer: 5' AGAGGTGCCCTCTCAAAA 3', Reverse (Rw): 5' TAGCTAGCACTCAGGAGGGTGA 3'. **C.** T7 endonuclease assay performed with the PCR product from three different iCas9 ES cell groups reveal the presence of two extra bands observed at around 470 bp and 300 bp only in the cells that was transduced with the virus and induced with DOX for 96 hrs. This indicates the mutations created in the *Trp53* gene of iCas9 mES cells that was transduced with the virus containing a plasmid with gRNA targeting *Trp53* and induced with doxycycline for 96 hrs

3.1.4 Evaluation of the iCas9 mouse model for tumor modeling

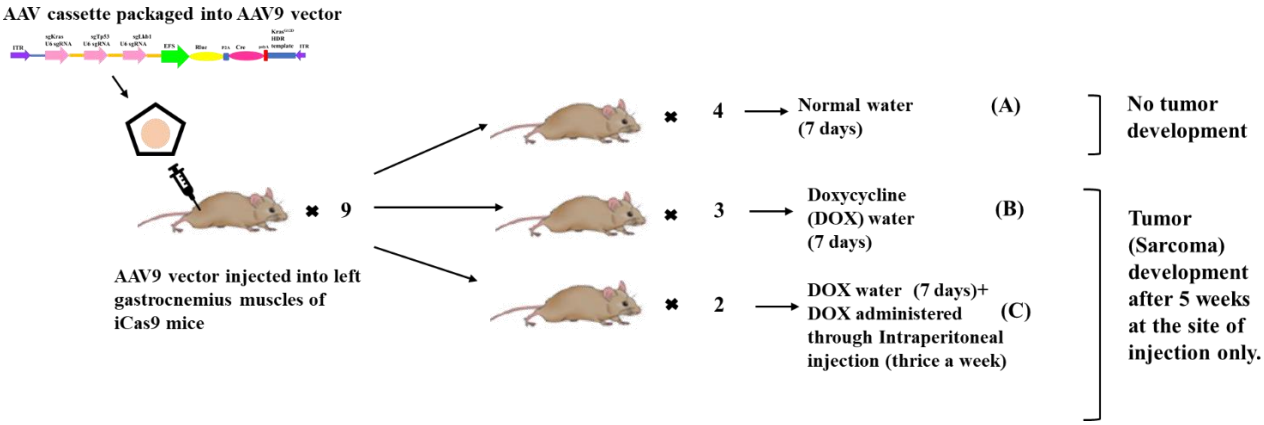
The iCas9 mouse had been previously used only for tracing the lineage history and the phylogeny of single hematopoietic stem cells[71]. However, this model has not yet been used to validate its ability to be used for sporadic tumor modeling.

We obtained an AAV vector containing sgRNAs targeting *Trp53*, *Lkb1*, *Kras* along with a homology arm for introducing G12D mutations into *Kras* from Addgene. The depicted AAV vector **(Figure 3.5 A)** was packaged into an AAV 9 serotype by Vector Builder®. The AAV9 vector was injected into the left gastrocnemius muscles of nine (N = 9) iCas9 mice to evaluate its ability to conjugate with Cas9 and create tumorigenic mutations ultimately leading to a tumor development from the soft tissues injected. The virus injected group was split into three different cohorts, (A) four animals were provided with normal drinking water, (B) three animals were provided with DOX in the drinking water for seven days, (C) two animals received DOX water and an additional bolus Intraperitoneal injection (IP) of DOX (3 X week) for seven days **(Figure 3.5 B).** Five weeks after AAV9 virus injection into the iCas9 mouse model, only the cohorts that received DOX water and/or supplemental bolus doses of DOX developed solid tumors characterized by a bulge at the site of injection whereas the group that received normal water did not develop any tumors **(Figure 3.5 C).** H and E staining analysis revealed the herringbone (fibroblasts cross aligned with each other) and a spindle shaped morphology of the tumor tissue suggesting that the tumor might be of fibroblastic origin[73] (fibrosarcoma) **(Figure 3.6 A).** Moreover, the two cell lines established also had spindle shaped fibroblastic features potentially indicating the mesenchymal- fibroblasts as the cell of origin of the tumors. **(Figure 3.6 B)** Sanger sequencing of the DNA from a cell line established from the tumor revealed anomalies in the *Lkb1* gene **(Figure 3.7 A)** compared to the sequencing reads from the wildtype muscle cell line established from a wild type Kh2iCas9 mouse **(Figure 3.7 B)**.

A.



B.



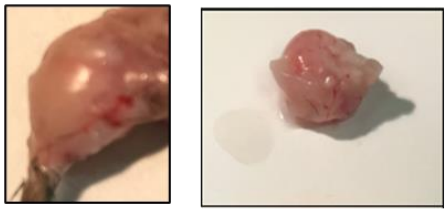
C.

Tissue collected from different iCas9 mouse groups after 5 weeks

i. **Normal Muscle tissue**

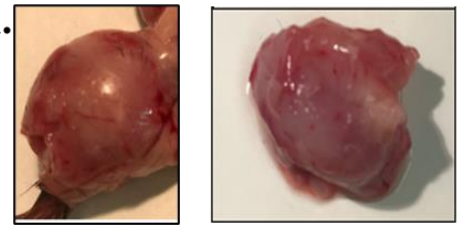
iCas9 mouse received AAV9 cassette and provided with normal water

ii. **Tumor mass**



iCas9 mouse received AAV9 cassette and provided with Doxycycline water for 7 days

iii. **Tumor mass**



iCas9 mouse received AAV9 cassette and provided with Doxycycline (DOX) water for 7 days and (3x) intraperitoneal DOX injection

Figure 3.5: iCas9 mice form tumors when injected with AAV9 cassette containing guides targeting *Trp53, Lkb1, Kras* and homology arms for introducing oncogenic *Kras* and induced with doxycycline

A. The AAV viral vector consists of inverted terminal repeats (ITR) at both of its terminal ends that allows for AAV virus to enter the host genome. 5' ITR is followed by U6 driven guide RNA targeting *Kras*, *Trp53* and *Lkb1*. Downstream of the guides is a EFS (elongation factor 1α short promoter) that drives luciferase reporter (r luc) and Cre recombinase. The luciferase and Cre recombinase are separated from each other by a P2A sequence. This sequence is followed by a homology arm for introducing the glycine (G) to aspartate (D) mutation in the 12^{th} amino acid that results in the oncogenic *Kras*. **B.** Schematic of the sporadic sarcoma modeling achieved by packaging the previously described AAV cassette into an AAV9 vector and injecting it into the left gastrocnemius muscle of the iCas9 mice (Number of mice (N) = 9). The injected iCas9 mice were split into three cohorts. Group A (N = 4) received only normal drinking water for 7 days. Group B (N = 3) received doxycycline water for seven days and Group C (N = 2) received doxycycline water plus intraperitoneal injection of doxycycline for three times a week. **C.** The panel in the left in all the three images illustrates the normal/tumor development in the injected region after five weeks. The sarcoma development in panel ii and iii is characterized by a predominant bulge around the injected hindlimbs. The panel in the right side depicts the resected tumors from the cohorts that received both doxycycline (ii) and intraperitoneal dose of doxycycline (iii). Larger size tumors were observed in the cohort (iii) that received both doxycycline administered water and also intraperitoneal dose of doxycycline indicating that the induction of Cas9 is increased by higher doses of doxycycline. (Total number of animals (N): 9)

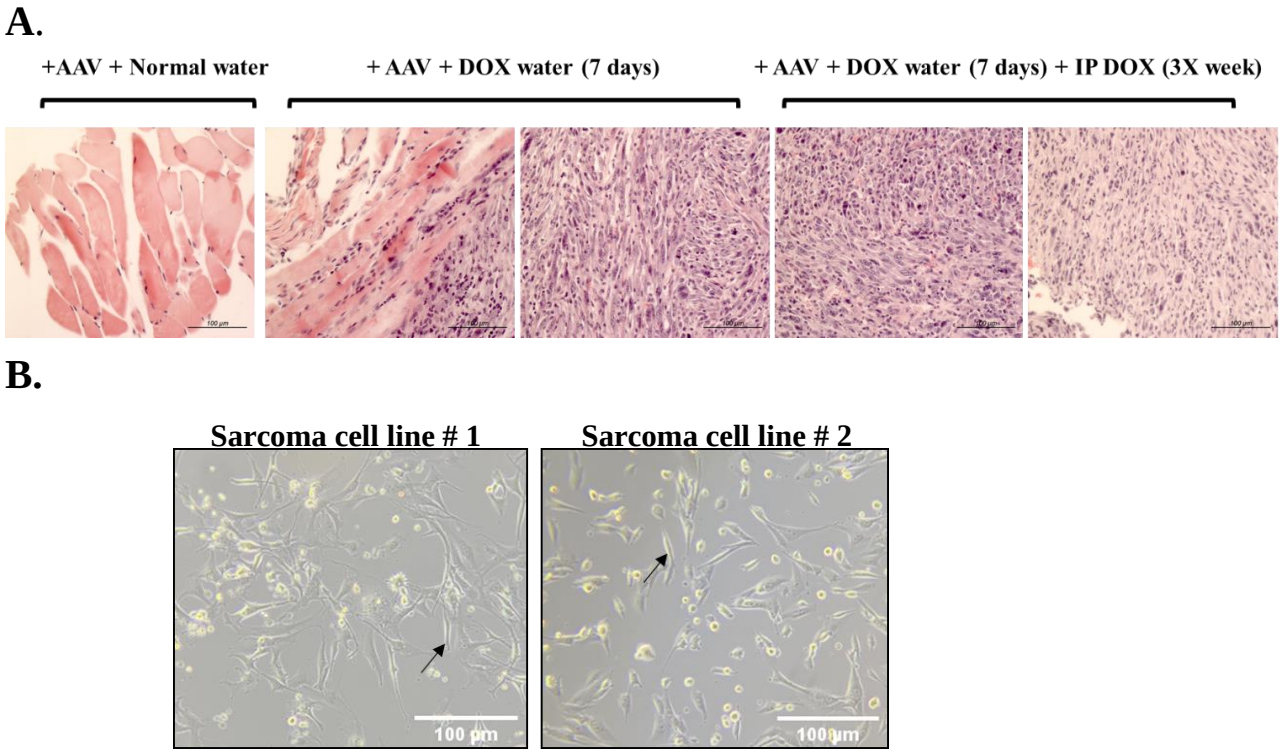
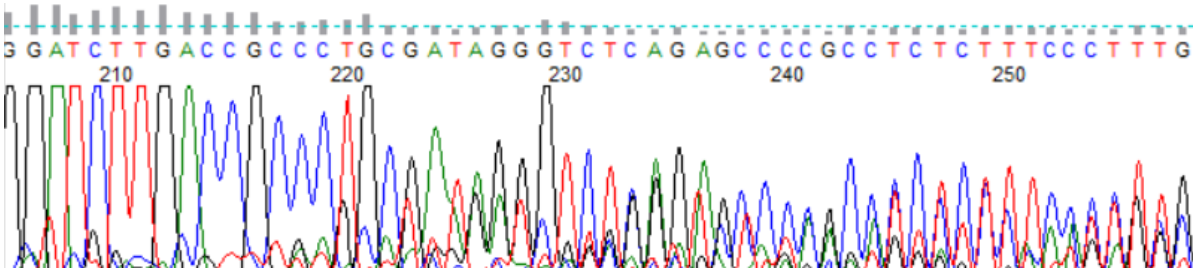


Figure 3.6: Hematoxylin and Eosin staining of the tumor tissues and the cell lines established from them reveals the herringbone morphology and fibroblastic phenotype suggesting that the tumor may be a fibrosarcoma

A. Hematoxylin and Eosin staining analysis of the tissue harvested from the AAV9 injected iCas9 mice that were administered with the normal drinking water (+AAV + Normal water), administered with DOX water for 7 days (+AAV+ DOX), administered with DOX water for 7 days and intraperitoneal injection of DOX thrice a week (+AAV+ DOX+ IP DOX). It could be observed that the tumor tissues have a herringbone morphology wherein the fibroblasts are in perpendicular to each other typically associated with a fibrosarcoma phenotype (yellow arrows in A). The scale bar represents 100 µm. **B.** Two cell lines established from the tumors resected from both +DOX and ++DOX conditions had a typical fibroblastic morphology (arrows) under a bright field microscope further substantiating the mesenchymal-fibroblasts as the cell of origin of the tumors. The scale bar represents 100 µm

A. Sanger sequencing reads: Sarcoma cell line - *Lkb1*

Forward sequencing primer: GAGGTGACGCTCAGGGATAGAT



B. Sanger sequencing reads: Wild type Muscle cell line - *Lkb1*

Forward sequencing primer: GAGGTGACGCTCAGGGATAGAT

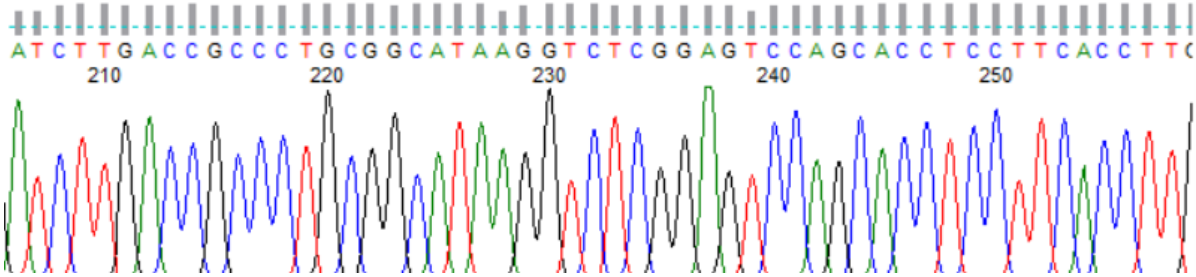


Figure 3.7: Sanger sequencing depicts the irregular reads in the *Lkb1* gene in a cell line established from the tumor compared to the wild type muscle cell line

A small portion of the sanger sequencing read from the guide RNA targeted region of the *Lkb1* gene of the sarcoma cell line (B) depicts great variability in the sequencing reads compared to the same region from the wild type muscle cell line.

PART 2: 3.2 To develop a mouse line with the conditional and inducible (COIN) expression of wt/dCas9 and the MPH gene activation complex

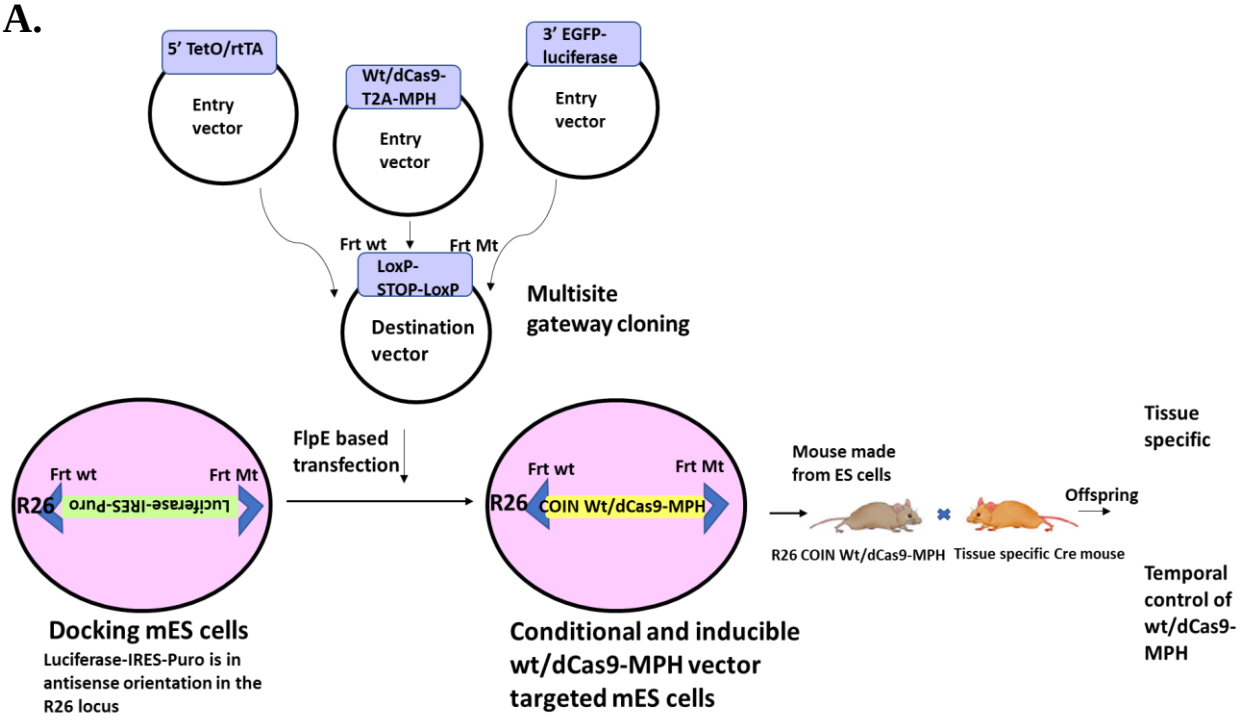
3.2.1 Background: As demonstrated in Part1 of this thesis, the iCas9 mouse model is a novel and powerful tool for modeling aspects of tumorigenesis. However, tumorigenesis is a complex process that also involves upregulation of oncogenes and downregulation of tumor suppressors at the same time in certain cases. Previous overexpression mouse models developed through overexpression of a transgene using artificial promoters and random transgene integration often does not recapitulate pathophysiological relevant gene expression levels[63]. Also, developing a transgenic mouse model with both activating and inactivating mutations is very tedious and time consuming and requires breeding of several conditional gain and loss of function alleles together with a tissue-specific Cre line. In order to achieve such endogenous gene activation and knock outs in the iCas9 mouse model, the fusion proteins and DgRNAs that would allow for transgene activation would have to be separately delivered into the desired cell type or tissue. However, this process may be difficult owing to the size of the fusions to be virally delivered or in case of introducing them into difficult to reach organs such as the heart or thymus.

We therefore wanted to create a mouse line wherein the Cas9 and a gene activation complex (MPH) are already embedded in the mouse at the well-defined safe harbor ROSA26 locus and their expression could be controlled in a tissue specific and temporal manner. We chose to develop the aforementioned mouse model based upon the Haigh lab's well established ES cell based transgenesis protocols[26,27]. Conditional and inducible (COIN) variants of wt/dCas9-MPH components would be knocked into parental G4 ES cells[26] that carry a docking site previously engineered into the R26 locus through recombinase-mediated cassette exchange (RMCE)[26,27] **(Figure 3.8 A)**. The COIN wt/dCas9-MPH ES cells once characterized at an *in vitro* level would then be used to develop a mouse line.

The mouse made from the COIN wt/dCas9-MPH mES cells through diploid aggregation technologies would have the potential to express both Cas9 and MPH in a conditional and inducible manner in all its adult tissues **(Figure 3.8 A)**. Breeding this particular model with tissue specific Cre lines would result in the generation of a mouse line wherein the Cas9 and the MPH complex would be expressed in the specific tissue of interest and only with the induction of DOX through drinking water. Delivering the desired guide RNAs that target different genes would allow for upregulated (MS2 based dgRNAs) and/or downregulated (gRNAs) gene expression of numerous genes only in that particular tissue of interest where Cas9 and MPH are expressed upon DOX induction **(see Figure 3.8 B).** In this way, sporadic tumor modeling involving upregulation of oncogenes could be effectively achieved from

tissues belonging to any of the three germ layers just by transducing the desired tissue with a combination of different guide RNAs targeting the promoter-regions of different genes and allowing for assembly of Cas9/dgRNA/MPH transcriptional activation complex at the promoter of these genes. Compared to the conventional gene knock outs or transgene overexpression mice models, the Cas9 based gene modulatory system allows researchers to manipulate the levels of endogenous gene expression programs and can obtain endogenous gene expression levels that are observed in pathological conditions such as cancer. Aspects of tumorigenesis such as oncogene addition (dependence of tumor survival and proliferation on the sole expression of an oncogene and its associated signaling[64]) could be precisely studied in this model owing to the DOX mediated control that allows for the modulation of Cas9 mediated gene expression changes. Moreover, this type of *in vivo* gene modulating system would facilitate the interrogation of numerous signaling pathways by simultaneously activating or knocking out different effectors involved in them which is difficult to recapitulate in conventional transgenic mouse models. This mouse model would also serve as a useful *in vivo* platform for genome wide activation and inactivation screens.

A.


5' TetO/rtTA
Entry vector
Wt/dCas9-T2A-MPH
Entry vector
3' EGFP-luciferase
Entry vector
Frt wt
Frt Mt
LoxP-STOP-LoxP
Destination vector
Multisite gateway cloning
Frt wt
Frt Mt
R26
Luciferase-IRES-Puro
FlpE based transfection
Frt wt
Frt Mt
R26
COIN Wt/dCas9-MPH
Mouse made from ES cells
Offspring
R26 COIN Wt/dCas9-MPH
Tissue specific Cre mouse
Tissue specific
Temporal control of wt/dCas9-MPH
Docking mES cells
Luciferase-IRES-Puro is in antisense orientation in the R26 locus
Conditional and inducible wt/dCas9-MPH vector targeted mES cells

B.

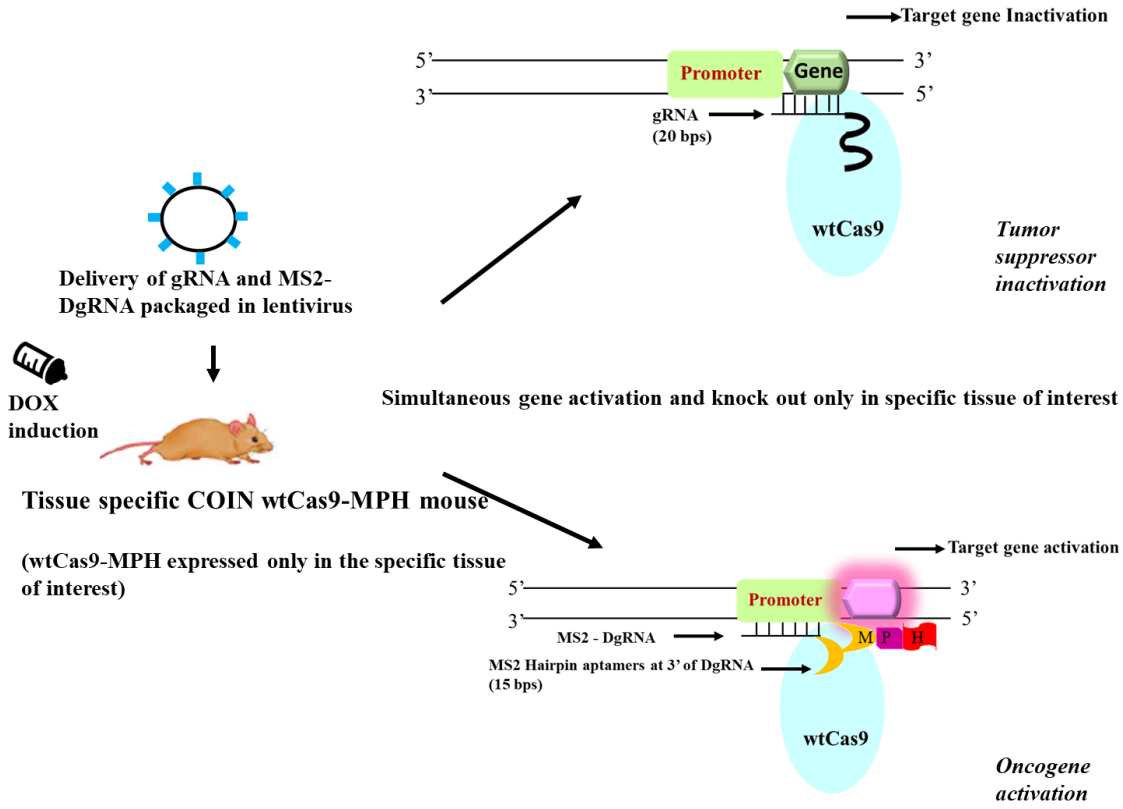

Target gene Inactivation
5'
3'
Promoter
Gene
3'
5'
gRNA
(20 bps)
wtCas9
Tumor suppressor inactivation
Delivery of gRNA and MS2-DgRNA packaged in lentivirus
DOX induction
Simultaneous gene activation and knock out only in specific tissue of interest
Tissue specific COIN wtCas9-MPH mouse
(wtCas9-MPH expressed only in the specific tissue of interest)
Target gene activation
5'
3'
Promoter
3'
5'
MS2 - DgRNA
M P H
MS2 Hairpin aptamers at 3' of DgRNA
(15 bps)
wtCas9
Oncogene activation

Figure 3.8: Schematic overview of the development and utility of the conditional and inducible (COIN) Cas9-MPH mouse and its utility in simultaneous gene activation and gene inactivation

A. Generation of COIN wt/dCas9-MPH mouse. The COIN wt/dCas9-MPH vector generated through Gateway cloning when targeted into the Haigh lab's docking mES cells allows to generate an ES cell line with the (COIN) conditional and inducible expression of Cas9 and MPH gene activation complex. The COIN wtCas9-MPH mES cell line when made into a mouse through diploid aggregation technologies wherein the ES cells are combined with a morula stage embryo result in the generation of a chimeric offspring expressing wt/dCas9-MPH variants in all its adult tissues in a conditional and inducible manner. **B. Utility of the COIN wt/dCas9-MPH mouse model.** Breeding the conditional and inducible (COIN) wtCas9-MPH mouse model with a tissue specific Cre mouse results in the generation of an offspring expressing wtCas9-MPH only in the specific tissue of interest and only upon the administration of doxycycline through drinking water. Upon delivery of both normal length guide RNAs (around 20 bps) and MS2-DgRNAs (13-15 bp guides conjugated with MS2 hairpin loops) allows to perform simultaneous gene knock down and gene activation. The MS2-DgRNAs recruits the MPH complex and Cas9 to the promoter sequence of the desired gene of interest (eg. oncogene) and upregulates its gene expression levels. The gRNA recruits the Cas9 to the target gene (eg. Tumor suppressor) for creating double stranded breaks leading to gene knock out. This allows to model aspects of tumorigenesis requiring simultaneous gene activation and gene knock down. Both the gRNA and the MS2-DgRNA could be delivered in a single packaging vector for achieving simultaneous gene activation and knock out.

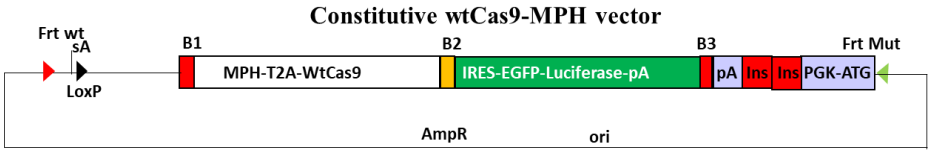
3.2.2 Development and characterization of positive control mES cell lines

Before we developed our COIN wtCas9-MPH mES system, we created and characterized three positive control mES cell systems for validating different aspects of conditional and inducible wt/dCas9-MPH mES cell system.

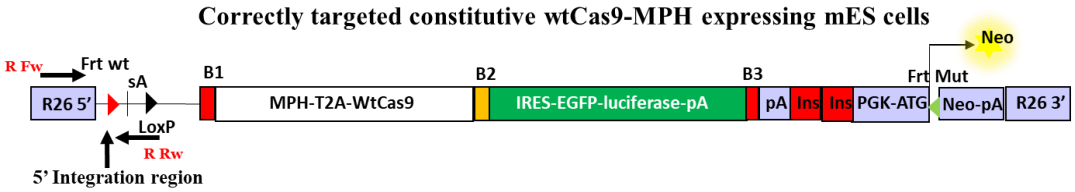
3.2.2.1 Development of positive control mES cell lines for the COIN wtCas9-MPH system

The first mES cell line did not rely upon the Cre-mediated excision of a transcriptional stop cassette or DOX induced expression in order to express Cas9 and MPH. These cells are expected to constitutively express wt/dCas9-MPH and EGFP-luciferase from the endogenous R26 promoter and are simple denoted as constitutive wt/dCas9-MPH expressing cells. The constitutive wtCas9-MPH gene targeting vector **(Figure 3.9 A)** previously developed vector through Gateway cloning was targeted into the docking mES cells through FlpE based cassette exchange reaction **(Figure 3.9 B).** The correctly targeted neomycin resistant clones were validated through PCR using a pair of primers indicated in the figure as R Fw and R Rev that specifically amplifies the 5' integration region of the vector into the R26 locus of docking mES cells. The amplified region is seen at around 560 bp when the PCR products are run on an 1.5% agarose gel. **(Figure 3.9 C).** The second cell line is the inducible wtCas9-MPH mES cells that is expected to express Cas9-MPH and EGFP-luciferase when just induced with DOX. In order to create this cell line, the destination vector used for developing the COIN wt/dCas9-MPH vector was incubated with a Cre recombinant protein for 1 hr at 37ºC and then briefly incubated at 70ºC to inactivate the Cre protein. This Cre excised destination was used in a similar multisite Gateway cloning with 5' rtTA/Tet on, wt/dCas9-MPH and 3' EGFP-luciferase vectors to generate a Cre excised inducible wt/dCas9-MPH gene targeting vector which was **(Figure 3.9 D)** validated for correct size through restriction digestion with Pac1. Restriction digestion with Pac1 enzyme depicts the expected cleavage of the inducible wt/dCas9-MPH vectors with two bands observed at around 14 KB and 4 KB respectively **(Figure 3.9 E)**. The vector was targeted into the docking mES cells and correctly targeted neomycin resistant clones **(Figure 3.9 F)** were validated through PCR using a pair of primers indicated in the figure as R Fw and R Rev that specifically amplifies the 5' integration region of the vector into the R26 locus of docking mES cells. The amplified region is seen at around 560 bp when the PCR products are run on an 1.5% agarose gel. **(Figure 3.9 G).**

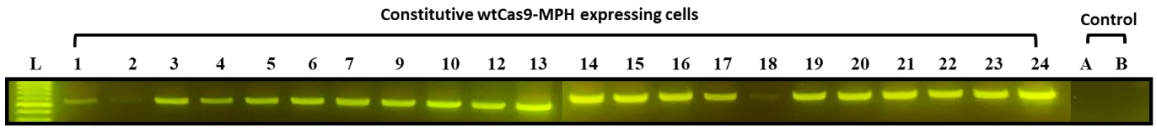
A.



B.



C. 5' Integration validation of the correctly targeted constitutive wtCas9-MPH vector containing mES cells

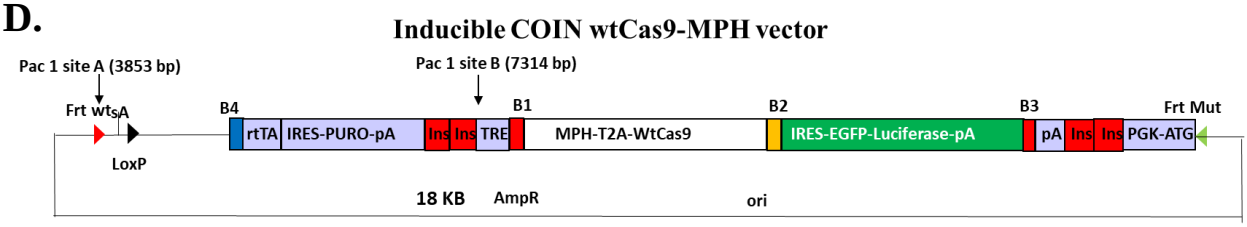


1-24 ➡ Correctly targeted Constitutive wtCas9-MPH mES cells

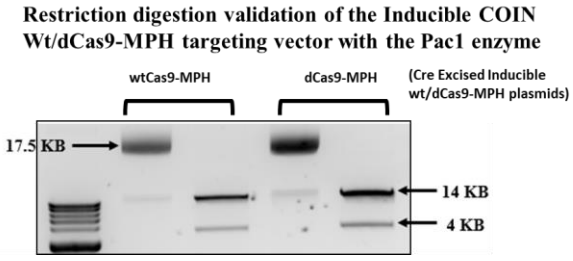
A ➡ Docking mES cells

B ➡ ddH20

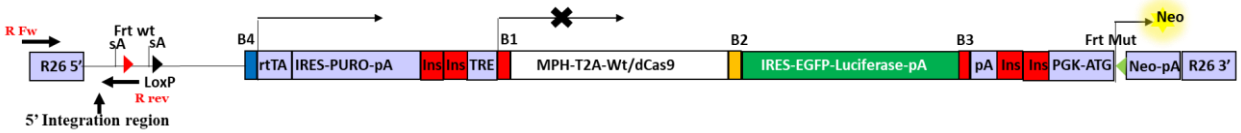
D.



E.

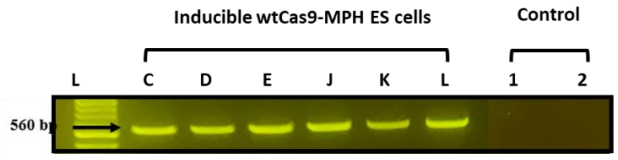


F.



G.

5' Integration validation of the correctly targeted Inducible COIN wtCas9-MPH vector containing mES cells



C – L ➡ Inducible wtCas9-MPH ES cells

1 ➡ Docking cells

2 ➡ ddH20

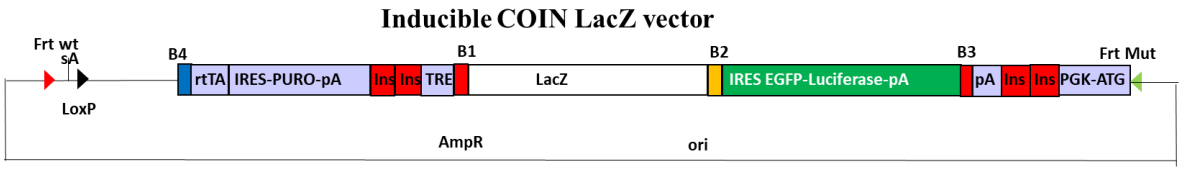
Figure 3.9: Generation of constitutive wtCas9-MPH mES cell line and inducible wtCas9-MPH mES cell line

A. The constitutive wtCas9-MPH vector consists of a wt FlpE recombination site (Frt) at its 5' end followed by a splice acceptor site (sA) and a loxP site. This sequence is followed by a B1 recombination site and a MPH-T2A-WtCas9 sequence and a B2 Gateway cloning recombination site. This is followed by a IRES-EGFP-luciferase-pA sequence. This cassette is followed by a B3 Gateway cloning recombination site and 2X Insulator sequence (Ins) and a PGK-promoter (3-phosphoglycerate kinase) and a ATG start codon followed by a Frt Mutant site. **B.** The correctly targeted constitutive wtCas9-MPH mES cells are expected to express MPH-wtCas9 and EGFP-Luciferase from the R26 promoter in a constitutive manner i.e., no Cre excision or any induction is required. **C.** The correctly targeted Neomycin resistant clones are validated through PCR using a pair of primers indicated in the figure as R Fw and Rev that specifically amplifies the 5' integration region of the vector into the R26 locus of docking mES cells. The amplified region is seen at around 560 bp when the PCR products are run on an 1.5% agarose gel. **D.** The inducible COIN wt/dCas9-MPH vector consists of a wt FlpE recombination site (Frt) at its 5' end followed by a splice acceptor site and a loxP site. This sequence is followed by a B4 Gateway cloning recombination site that is followed by a rtTA-IRES-Puromycin-pA sequence. This cassette is followed by 2X Insulator sequence (Ins) and a TRE promoter. The TRE promoter is followed by a B1 recombination site and a MPH-T2A-Wt/dCas9 sequence and a B2 Gateway cloning recombination site. This is followed by a IRES-EGFP-luciferase-pA sequence. This cassette is followed by a B3 Gateway cloning recombination site and 2X Insulator sequence (Ins) and a PGK-promoter (3-phosphoglycerate kinase) and a ATG start codon followed by a Frt Mutant site. Cre excised inducible wtCas9-MPH gene targeting vector. **E.** Restriction digestion validation of the inducible wtCas9-MPH gene targeting vector with the Pac1 enzyme. Two different bands are observed in the correctly digested vectors at around 14 KB and 4 KB when run on an 1.5% agarose gel. **F.** Correctly targeted inducible wtCas9-MPH expressing mES cells that are expected to express wtCas0-MPH and EGFP-luciferase when induced with doxycycline. **G.** The correctly targeted Neomycin resistant clones are validated through PCR using a pair of primers indicated in the figure as R Fw and R Rev that specifically amplifies the 5' integration region of the vector into the R26 locus of docking mES cells. The amplified region is seen at around 560 bp when the PCR products are run on an 1.5% agarose gel.

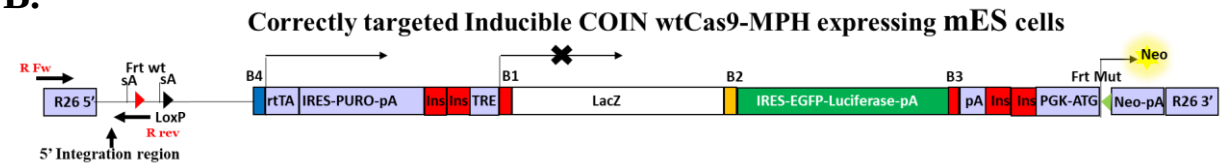
3.2.2.2 Development of positive control LacZ mES cell lines for DOX induction experiments

The third control inducible LacZ mES cell line was developed for detecting the inducibility of TRE promoter by the doxycycline drug (DOX). This cell line was created by targeting a previously published inducible LacZ and EGFP-luciferase expressing vector **(Figure 3.10 A)** into the docking mES cells and correctly targeted neomycin resistant clones **(Figure 3.10 B)** were validated through PCR using a pair of primers indicated in the figure as R Fw and R Rev that specifically amplifies the 5' integration region of the vector into the R26 locus of docking mES cells. The amplified region is seen at around 560 bp when the PCR products are run on an 1.5% agarose gel **(Figure 3.10 C).** EGFP analysis through flow cytometry on DOX induced and non-induced cells revealed that around 20% of cells expressed EGFP only in DOX treated cells and not in the non-induced cells**.** X-gal staining on DOX treated clones revealed LacZ staining characterized by the cells turning green-blue only in the DOX treated clones thereby substantiating the inducibility of the DOX to turn on the TRE promoter to drive the expression of LacZ **(Figure 3.10 D).**

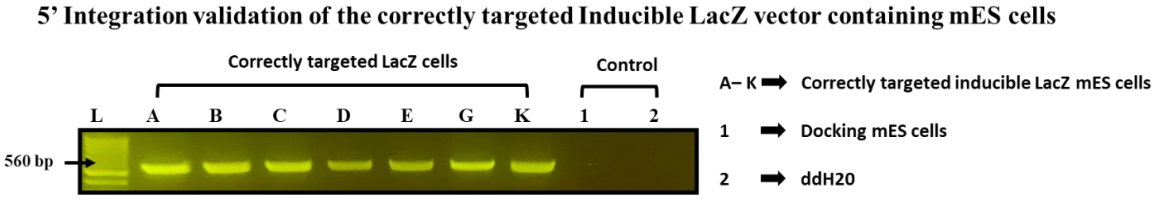
A.


Inducible COIN LacZ vector
Frt wt
sA
LoxP
B4
rtTA
IRES-PURO-pA
Ins
Ins
TRE
B1
LacZ
B2
IRES EGFP-Luciferase-pA
B3
pA
Ins
Ins
PGK-ATG
Frt Mut
AmpR
ori

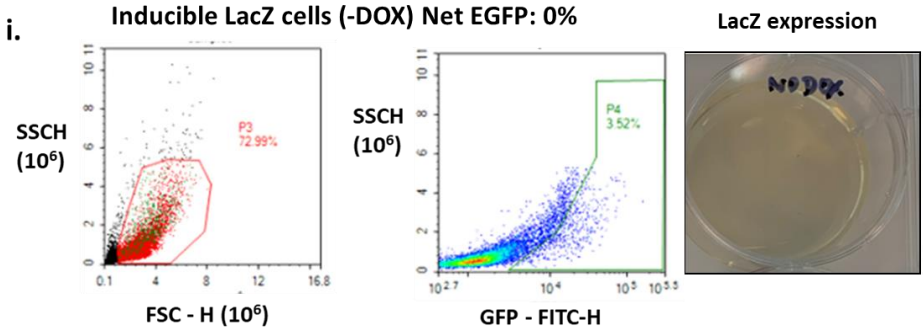
B.


Correctly targeted Inducible COIN wtCas9-MPH expressing mES cells
R Fw
R26 5'
Frt wt
sA
sA
LoxP
R rev
5' Integration region
B4
rtTA
IRES-PURO-pA
Ins
Ins
TRE
B1
LacZ
B2
IRES-EGFP-Luciferase-pA
B3
pA
Ins
Ins
PGK-ATG
Frt Mut
Neo
Neo-pA
R26 3'

C.


5' Integration validation of the correctly targeted Inducible LacZ vector containing mES cells
Correctly targeted LacZ cells
Control
L
A
B
C
D
E
G
K
1
2
560 bp
A– K ➡ Correctly targeted inducible LacZ mES cells
1 ➡ Docking mES cells
2 ➡ ddH20

D.


i.
Inducible LacZ cells (-DOX) Net EGFP: 0%
LacZ expression
SSCH (10⁶)
P3
72.99%
FSC - H (10⁶)
SSCH (10⁶)
P4
3.52%
GFP - FITC-H
NoDox

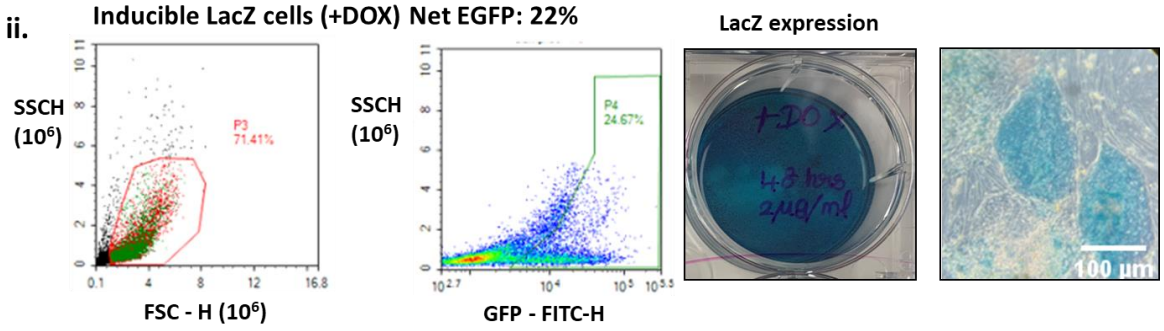

ii.
Inducible LacZ cells (+DOX) Net EGFP: 22%
LacZ expression
SSCH (10⁶)
P3
71.41%
FSC - H (10⁶)
SSCH (10⁶)
P4
24.67%
GFP - FITC-H
+DOX
100 µm

Figure 3.10: Generation and validation of inducible LacZ expressing mES cell line

A. The inducible COIN LacZ vector consists of a wt FlpE recombination site (Frt) at its 5' end followed by a splice acceptor site and a loxP site. This sequence is followed by a B4 Gateway cloning recombination site that is followed by a rtTA-IRES-Puromycin-pA sequence. This cassette is followed by 2X Insulator sequence (Ins) and a TRE promoter. The TRE promoter is followed by a B1 recombination site and a LacZ sequence and a B2 Gateway cloning recombination site. This is followed by a IRES-EGFP-luciferase-pA sequence. This cassette is followed by a B3 Gateway cloning recombination site and 2X Insulator sequence (Ins) and a PGK-promoter (3-phosphoglycerate kinase) and a ATG start codon followed by a Frt Mutant site. **B**. The Correctly targeted inducible LacZ expressing mES cells are expected to express LacZ and EGFP-luciferase when induced with doxycycline. **C.** The correctly targeted Neomycin resistant clones are validated through PCR using a pair of primers indicated in the figure as R Fw and R Rev that specifically amplifies the 5' integration region of the vector into the R26 locus of docking mES cells. The amplified region is seen at around 560 bp when the PCR products are run on an 1.5% agarose gel. **D.** Forward scatter (FSC-H) vs side scatter (SSC-H) plot is used to cover the region of interest in the inducible LacZ expressing cell populations (the left most panels of the Figure D). Cells expressing EGFP were profiled through FITC channel. This revealed that around 22% of cells expressed EGFP in response to DOX stimulation and are represented in the green quadrant in the SSC-H vs GFP-FITC plot in the inducible COIN LacZ cells treated with DOX. The non-induced (-DOX) cells do not express any EGFP levels. Similarly, X-Gal staining of the DOX treated and non-DOX treated inducible LacZ expressing mES cell clone revealed LacZ expression only in the DOX treated clones. The scale bar represents 100 μm.

3.2.2.3 Flow cytometric characterization of the three positive control mES cell lines

With the success of DOX induction obtained from the inducible LacZ-EGFP expressing clones, the other positive control mES cells clones (Constitutive wtCas9-MPH and Inducible wtCas9-MPH mES cell clones) were evaluated for EGFP expression through flow cytometry. DOX induction was used simultaneously for the LacZ EGFP-Clone as well as the inducible wtCas9-MPH clone. Flow cytometry analysis revealed that 55% of constitutive wtCas9-MPH expressing mES cell lines expressed the highest levels of net EGFP expression suggesting that Cas9-MPH is also expressed from the R26 promoter as they are the same part of the bicistronic vector **(Figure 3.11 B)**. 22% of inducible LacZ expressing cells expressed EGFP when administered with doxycycline **(Figure 3.11 C).** The single clone of inducible wtCas9-MPH mES cells analyzed under both DOX induced and non-induced conditions had around 5% of cells expressing EGFP suggesting they may also be expressing wtCas9-MPH **(Figure 3.11 D).** Autofluorescence levels were eliminated by using the docking mES cells for setting the gates **(Figure 3.11 A).**

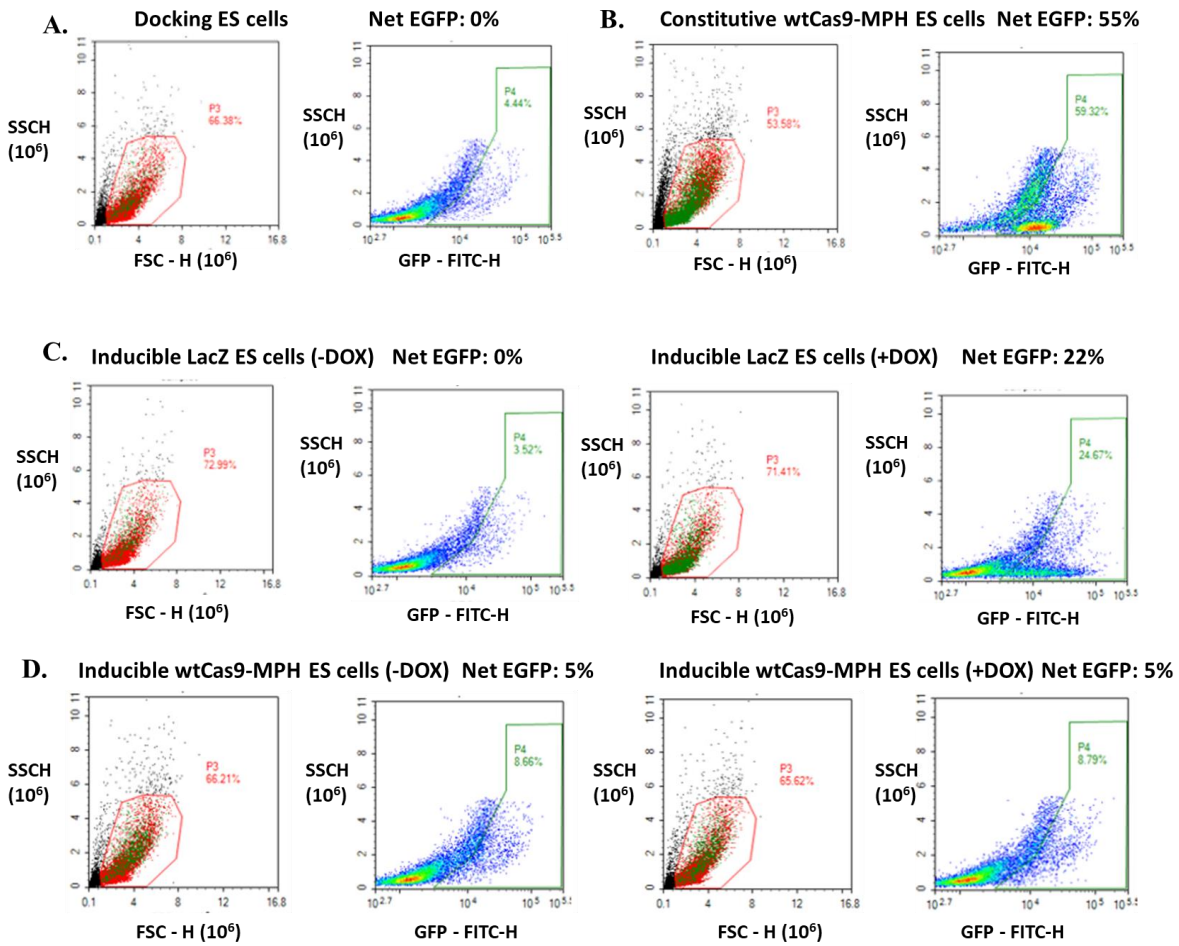


Figure 3.11: Flow cytometer profiling reveals that large levels of constitutive wtCas9-MPH and DOX induced LacZ mES cells express EGFP

Forward scatter (FSC-H) vs side scatter (SSC-H) plot is used to cover the region of interest in different mES cell populations. Cells expressing EGFP were profiled through FITC channel. **A.** Lack of EGFP expression in parental docking ES cells. **B.** Around 55% of cells express EGFP in constitutive wtCas9-MPH expressing mES cells. **C.** Around 22% of inducible LacZ expressing cells treated with DOX express EGFP whereas no net EGFP expressing cells were observed in non-DOX treated clones. **D.** Around 5% of cells express EGFP in inducible wtCas9-MPH expressing cells treated w/without DOX.

3.2.2.4 Functional characterization of constitutive wtCas9-MPH expressing cells

Owing to a high proportion of cells expressing EGFP in the constitutive wtCas9-MPH clones, we postulated that these cells would be expressing wtCas9-MPH as they are part of the same polycistronic cassette. In order to validate if the wtCas9-MPH complex is split into functional wtCas9 and MPH, a plasmid containing a guide RNA targeting *Trp53* was packaged into a lentiviral vector by our lab Research Associate Dr. Aissa Benyoucef and transduced into them. Fluorescence microscope imaging revealed that only the transduced cells have BFP expression **(Figure 3.12 A)**. T7 endonuclease assay revealed the presence of two extra bands sized at around 470 bp and 300 bp in addition to the main band at 800 bp only in the constitutive wtCas9-MPH expressing cell line. This result has validated that the Constitutive wtCas9-MPH expressing cells have functional Cas9 capable of creating double stranded breaks at the target site **(Figure 3.12 B).**

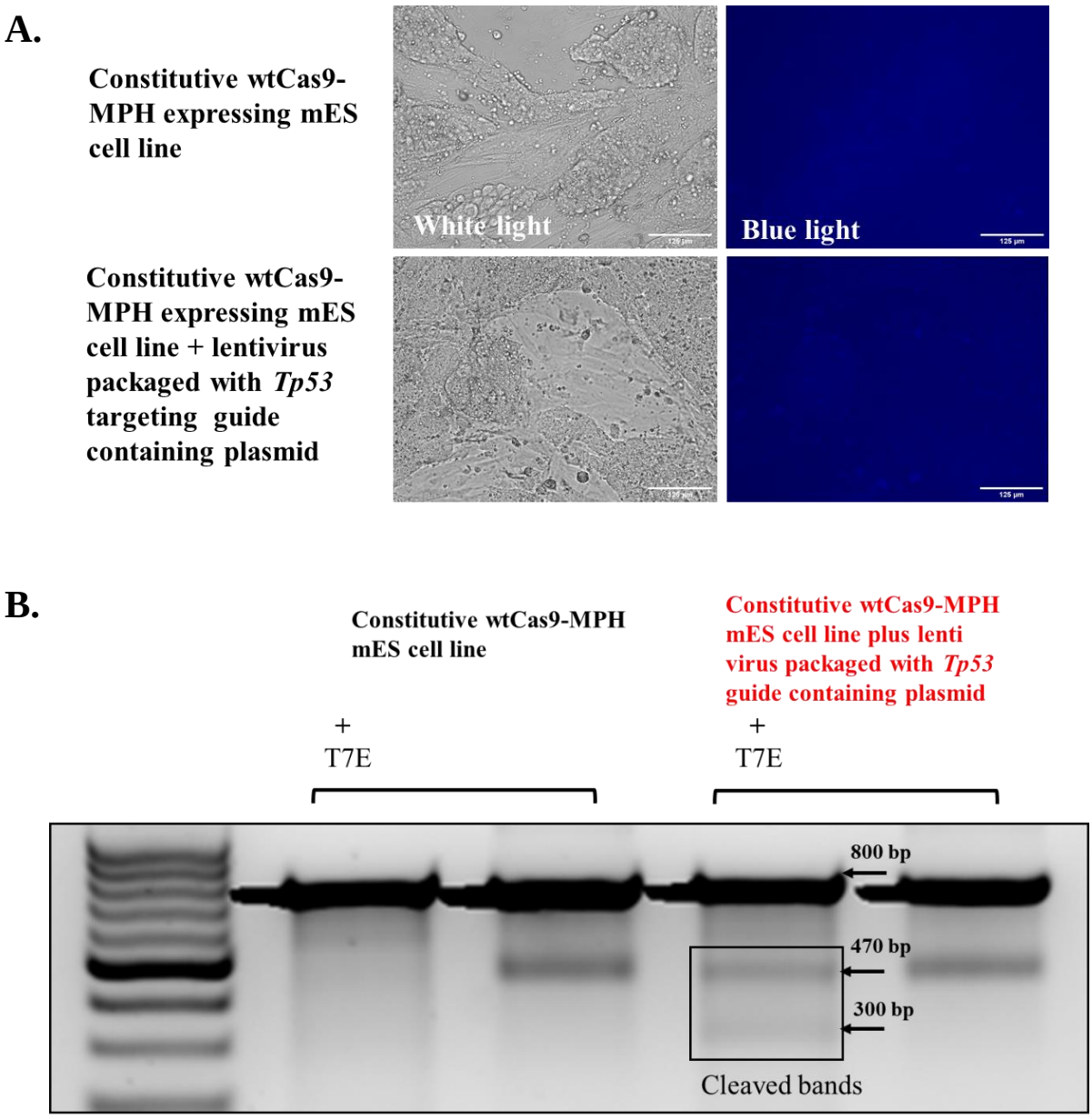


Figure 3.12: T7 endonuclease assay reveals the mutations created by Cas9-gRNA in the *Trp53* gene of the constitutive wtCas9-MPH mES cells transduced with a guide targeting *Trp53*

A. Fluorescence microscopy images of the constitutive wtCas9-MPH expressing ES cells transduced with/without a plasmid containing a guide targeting *Trp53*. BFP expression is observed in cells transduced with the lentivirus packaged with *Trp53* targeting guide RNA containing plasmid. The scale bar represents 125 μM **B.** T7 Endonuclease assay performed with the PCR product from two different constitutive wtCas9-MPH ES cell groups (control group and virus transduced groups) revealed the presence of extra bands sized at 470 bp and 300 bp only in the cells that was transduced with the virus. This indicated the mutation created in the *Trp53* gene of the constitutive wtCas9-MPH mES cell line that was transduced with the virus containing plasmid targeting *Trp53*.

3.2.3 Generation and characterization of the conditional and inducible (COIN) wt/dCas9-MPH expressing mES cell line

The mES cell line expressing COIN wtCas9-MPH complex was developed by creating a *in cis* COIN wtCas9-MPH vector and targeting it to our lab's docking mES cell line.

3.2.3.1 Development of *in cis* conditional and inducible (COIN) wt/dCas9-MPH expressing mES cell line at the R26 locus

A gene targeting plasmid vector that would express *in cis* (on the same allele) a COIN wt/dCas9-MPH sequence was first developed through the efficient Gateway cloning strategies previously developed in our lab[26]. An overall schematic of the whole process is provided in **Figure 3.13.** The three multi-site Gateway cloning compatible plasmid entry vectors (5' rtTA/Tet on vector, MPH-T2A wt/dCas9 vector, 3' EGFP-luciferase) and the destination vector were utilized for generating the targeting vector. The 5' Tet on vector is flanked with L4 and R1 Gateway cloning sites and consists of revere tetracycline controlled transactivator (rtTA) followed by an Internal ribosome entry sequence (IRES)-Puromycin-pA cassette. The internal ribosome entry sequence allows for the transcription of two or more genes from one single promoter[29]. This part of the vector is followed by 2X chicken beta globin insulator core sequences (Ins) and an inducible tetracycline (Tet)- responsive element (TRE) which is a multimerized rtTA binding sites and a minimal CMV promoter. The incorporation of insulator sequences has been previously shown by the Haigh lab to protect the CMV promoter from promoter interference and potential epigenetic silencing[27]. The second Entry vector is flanked with L1 and L2 Gateway cloning sites that consists of MPH gene activation complex followed by a self-cleaving peptide T2A and wt/dCas9 variants. T2A peptide site allows two separate proteins to be generated from the fused protein by endogenous proteases[69]. The third Entry vector is flanked with R2 and L3 sites and consists of an IRES-EGFP-Luciferase-pA+ sequence for monitoring the expression of transgenes. The multisite Gateway compatible destination vector is flanked with two heterospecific FlpE recombination sites (Frt wt and Frt-F5 mutant sites) and consists of a neomycin-3XpA transcriptional STOP cassette flanked by loxP sites. Downstream, the CCDB cassette is present that is flanked with R4 and R3 Gateway cloning sites followed by an insulator sequence and a PGK promoter (3-phosphoglycerate kinase) and ATG start codon. The three Entry vectors were combined into one Destination vector in a single multisite Gateway cloning-based L-R reaction to create *in cis* conditional and inducible (COIN) wt/dCas9-MPH gene targeting vectors **(Figure 3.14 A)**. The LR reactions were transformed into DH5a electro-competent bacteria and the plasmids were amplified and verified for correct orientation and sequence through restriction digestion with Pac1 enzyme and Sanger sequencing. Restriction digestion with Pac1 restriction enzyme depicts the expected cleavage of the

COIN Wt/dCas9-MPH vectors with two bands observed at around 14 KB and 5 KB respectively. **(Figure 3.14 B).** The verified COIN wt/dCas9-MPH vectors were targeted into the R26 locus of our lab's docking mES cells which contains two heterospecific Frt sites and a neomycin gene without its promoter and a start codon through FlpE based lipofectamine transfection. The correctly targeted cells **(Figure 3.14 C)** were selected under G418 resistance (Neomycin resistance) that originates when the COIN wt/dCas9-MPH vector is correctly targeted into the R26 locus of the docking mES cells which restores the Neomycin promoter and the PGK start codon for the promoter less neomycin resistance gene. Individual neomycin resistant clones were picked after 12 days and screened for correct 5' integration of the vector into the docking site of the R26 locus by subjecting the extracted DNA to a PCR reaction using a pair of primer that specifically amplifies the integration site of the vector with the R26 locus of the docking mES cells. Upon subjecting the PCR reaction mixture to an 1.5% agarose gel electrophoresis, a band of around 560 bp is observed only in the correctly COIN wt/dCas9-MPH vector targeted mES cells **(Figure 3.14 D)**.

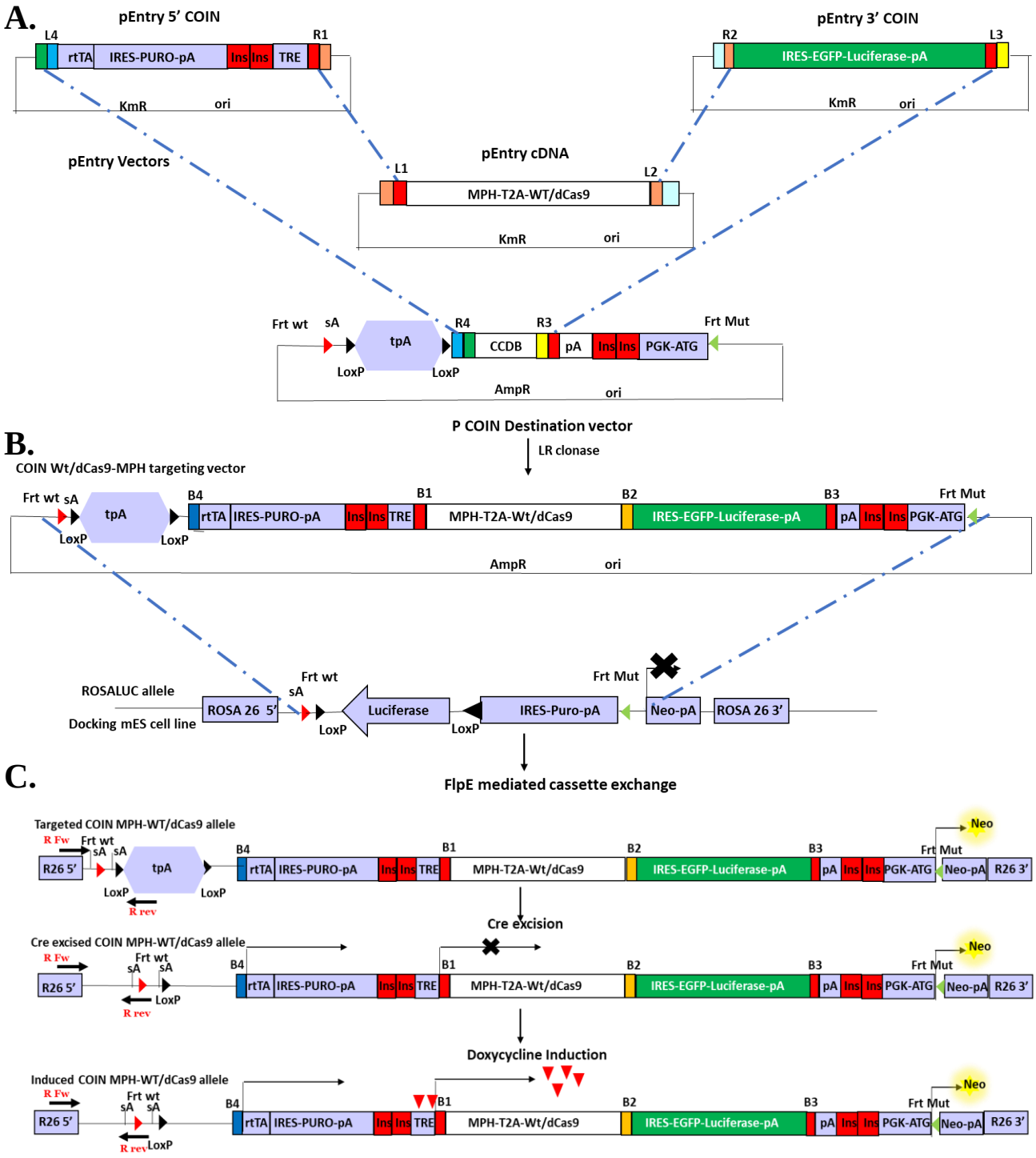
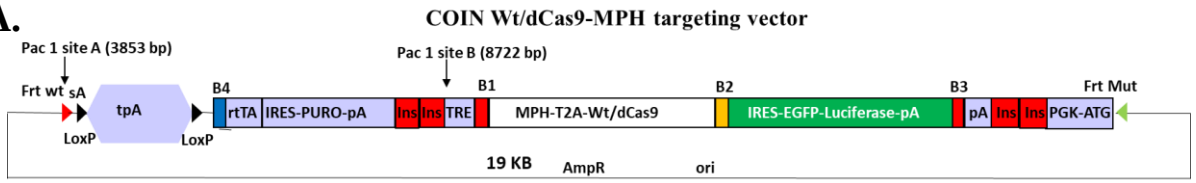

A.
pEntry 5' COIN
L4
rtTA
IRES-PURO-pA
Ins
Ins
TRE
R1
KmR
ori
pEntry 3' COIN
R2
IRES-EGFP-Luciferase-pA
L3
KmR
ori
pEntry Vectors
pEntry cDNA
L1
MPH-T2A-WT/dCas9
L2
KmR
ori
Frt wt
sA
tpA
LoxP
LoxP
R4
R3
CCDB
pA
Ins
Ins
PGK-ATG
Frt Mut
AmpR
ori
B.
P COIN Destination vector
LR clonase
COIN Wt/dCas9-MPH targeting vector
Frt wt
sA
tpA
LoxP
LoxP
B4
rtTA
IRES-PURO-pA
Ins
Ins
TRE
B1
MPH-T2A-Wt/dCas9
B2
IRES-EGFP-Luciferase-pA
B3
pA
Ins
Ins
PGK-ATG
Frt Mut
AmpR
ori
ROSALUC allele
Docking mES cell line
ROSA 26 5'
Frt wt
sA
LoxP
Luciferase
LoxP
IRES-Puro-pA
Frt Mut
Neo-pA
ROSA 26 3'
C.
FlpE mediated cassette exchange
Targeted COIN MPH-WT/dCas9 allele
R Fw
Frt wt
sA
sA
R26 5'
tpA
LoxP
LoxP
R rev
B4
rtTA
IRES-PURO-pA
Ins
Ins
TRE
B1
MPH-T2A-Wt/dCas9
B2
IRES-EGFP-Luciferase-pA
B3
pA
Ins
Ins
PGK-ATG
Frt Mut
Neo
Neo-pA
R26 3'
Cre excision
Cre excised COIN MPH-WT/dCas9 allele
R Fw
Frt wt
sA
sA
R26 5'
LoxP
R rev
B4
rtTA
IRES-PURO-pA
Ins
Ins
TRE
B1
MPH-T2A-Wt/dCas9
B2
IRES-EGFP-Luciferase-pA
B3
pA
Ins
Ins
PGK-ATG
Frt Mut
Neo
Neo-pA
R26 3'
Doxycycline Induction
Induced COIN MPH-WT/dCas9 allele
R Fw
Frt wt
sA
sA
R26 5'
R rev
LoxP
B4
rtTA
IRES-PURO-pA
Ins
Ins
TRE
B1
MPH-T2A-Wt/dCas9
B2
IRES-EGFP-Luciferase-pA
B3
pA
Ins
Ins
PGK-ATG
Frt Mut
Neo
Neo-pA
R26 3'

Figure 3.13: Detailed pictorial representation of generating and targeting the polycistronic conditional and inducible (COIN) wt/dCas9-MPH vectors into the docking mES cells

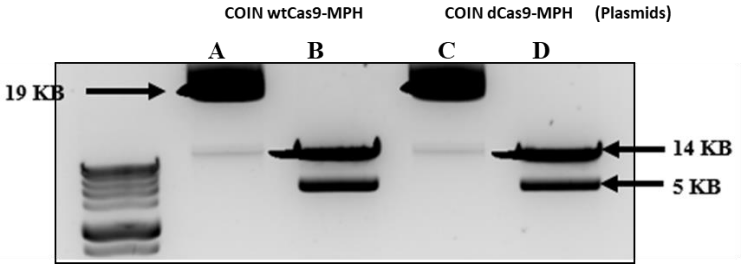
A. The three multi-site Gateway cloning compatible plasmid entry vectors and the destination vector were combined through LR reaction to generate the *in cis* conditional and inducible (COIN) MPH-wt/dCas9 gene targeting vectors. **B.** The COIN vector was then co transfected with a FlpE plasmid into the R26 locus of the GR4OSALUC cell line through trap mediated RMCE reaction which restores the neomycin resistance (Neo^R). **C.** Cre mediated removal of the STOP cassette results in the removal of the 3X pA-tpA+ STOP cassette and allows the expression of revere tetracycline controlled transactivator (rtTA) and the puromycin resistance gene as a bi-cistronic message from the endogenous R26 promoter. Inducible expression of MPH-wt/dCas9 can be achieved by administration with doxycycline which can also be monitored by the expression of EGFP. Correctly recombined clones could be screened by PCR (R F_w and R R_{Rev}) that confirms the amplification at 560 bp. (L1-4, R1-4, B1-4: Gateway compatible cloning sites, CCDB: control of cell death gene, SA: splice acceptor, IRES: internal ribosome entry sequence, pA: Polyadenylation sequence, Ins: Chicken beta globin insulator core sequence, TRE: Tet responsive element)

A.



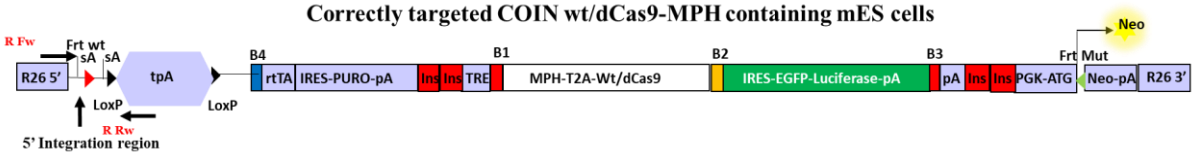
B.

Restriction digestion validation of COIN Wt/dCas9-MPH targeting vector using the Pac1 enzyme



C.

Correctly targeted COIN wt/dCas9-MPH containing mES cells



D.

5' Integration validation of the correctly targeted COIN wt/dCas9-MPH vector containing mES cells

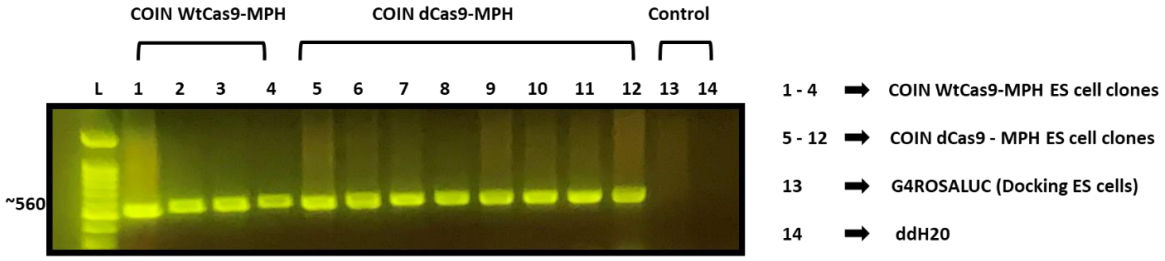


Figure 3.14: Conditional and inducible (COIN) wt/dCas9-MPH mES cell development

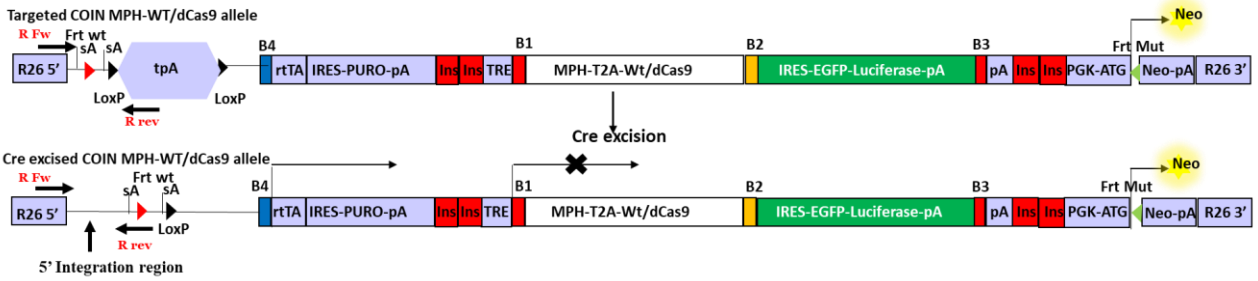
A. The conditional and inducible COIN wt/dCas9-MPH vector consists of a wt FlpE recombination site (Frt) at its 5' end followed by a splice acceptor site (sA) and loxP site flanked Neomycin resistance gene. This sequence is followed by a B4 Gateway cloning recombination site and a rtTA-IRES-Puromycin-pA sequence. This cassette is followed by 2X Insulator sequence (Ins) and a TRE (Tet responsive element) promoter. The TRE promoter is followed by a B1 recombination site and a MPH-T2A-Wt/dCas9 sequence and a B2 Gateway cloning recombination site. This is followed by a IRES-EGFP-luciferase-pA sequence. This cassette is followed by a B3 gateway cloning recombination site and 2X Insulator sequence (Ins) and a PGK-promoter (3-phosphoglycerate kinase) and a ATG start codon followed by a Frt Mutant site. **B.** Pac1 restriction enzyme sites is depicted in the previous figure. Restriction digestion validation of the COIN wt/dCas9-MPH vectors using the Pac1 restriction enzyme shows two different bands observed in the correctly digested vectors at around 14 KB and 5 KB when run on an 1.5% agarose gel. **C.** Correctly COIN wt/dCas9-MPH vector targeted neomycin resistant mES cell clones. **D.** The correctly targeted Neomycin resistant clones are validated through PCR using a pair of primers indicated in the figure as R Fw and Rev that specifically amplifies the 5' integration region of the vector into the R26 locus of docking mES cells. The amplified region is seen at around 560 bp when the PCR products are run on an 1.5% agarose gel.

3.2.3.2 Functional characterization of the COIN wtCas9-MPH mES cells

The COIN wt/dCas9-MPH mES cells when electroporated with a plasmid containing Cre recombinase and induced with DOX are expected to express wt/dCas9-MPH and EGFP-luciferase. Cre electroporation is expected to remove the transcriptional Neo STOP cassette and the resulting Cre excised COIN wt/dCas9-MPH cells would be expressing rtTA and puromycin resistance gene from the endogenous R26 promoter. The Cre excised rtTA expressing puromycin resistant cells when induced with DOX are expected to turn on the TRE promoter as DOX would form a complex with rtTA and translocate into the nucleus thereby activating it. The activated TRE promoter would in turn drive the expression of wt/dCas9-MPH and EGFP-luciferase.

During the course of these experiments, various COIN wt/dCas9-MPH mES cell clones were electroporated with a Cre plasmid to remove the transcriptional STOP cassette and to allow the expression of the targeted genes from the R26 promoter. It is expected that only the correctly Cre excised cells become resistant to the puromycin drug after Cre electroporation as it removes the transcriptional STOP cassette and allows the expression of rtTA-IRES-puromycin-pA+ resistance gene to be expressed from the R26 promoter **(Figure 3.15 A)**. However, even the non Cre excised cells were resistant to puromycin after 12 days of puromycin selection. Also, the putative Cre excised cells picked after 12 days of puromycin resistance selection lost their 5' integration band characterized by the absence of the band that was validated through PCR. **(Figure 3.15 B).** This factor thwarted these cells from being used for further analysis and is thoroughly discussed below in the subsequent discussion section.

A.



B.

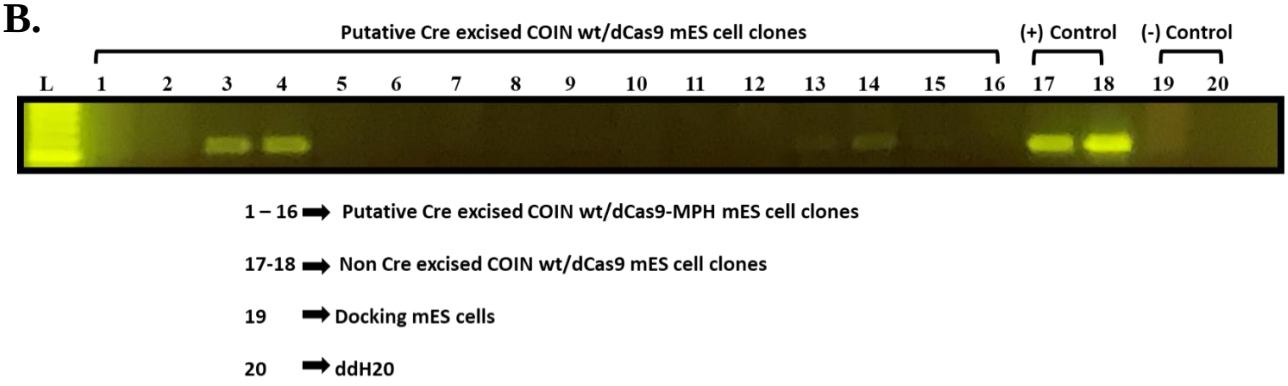


Figure 3.15: PCR analysis on putative Cre excised COIN wt/dCas9-MPH mES cell clones reveals the loss of 5' integration band

A. Electroporation of a Cre plasmid into the conditional and inducible (COIN) wt/dCas9-MPH clones is expected to remove the transcriptional neomycin STOP cassette and allow the expression of rtTA and puromycin resistance gene driven by the endogenous R26 promoter. Even after Cre electroporation, the cells should retain the 5' integration band. **B.** Numerous COIN wt/dCas9-MPH clones were electroporated with a Cre plasmid and puromycin resistant clones were subjected to 5' integration PCR which revealed the absence of the 5' integration band in them.

Chapter 4: Discussion

4.1 Inducible Cas9 mouse model – An efficient tool for tumor modeling

In the first Aim of this thesis, we have demonstrated that the inducible Cas9 mouse model is a novel resource for tumor modeling. Cas9 based genome editing is an effective strategy for disease modelling that reduces the dependence on conventional *in vivo* gene manipulation technologies which are both time consuming and expensive[57]. Previous studies have shown that loss of function mutations in Trp53 and oncogenic expression of Kras in skeletal muscles can lead to pleomorphic rhabdomyosarcoma formation[74]. Although studies have implicated the loss of function mutations in tumor suppressor Lkb1 in numerous cancer types such as lung cancer, pancreatic cancer and ovarian cancer, its role in sarcoma genesis remains elusive. This study investigated the synergy between the loss of function mutations in Trp53, Lkb1 and gain of function mutations in Kras created by Cas9 in initiating the tumor development in the normal gastrocnemius muscles of the iCas9 mouse model. Injection of a previously published AAV vector containing packaged into an AAV9 virus containing guides targeting *Trp53*, *Lkb1*, *Kras* and along with homology arms for introducing oncogenic *Kras*[20] into the gastrocnemius muscles of the inducible Cas9 mouse model along with the induction of DOX led to the formation of tumors within 5 weeks. The tumors were characterized by a prominent bulge over the hind legs where the mice received the injection and also administered with DOX through drinking water. Moreover, development of larger tumors in the cohorts that received both DOX in drinking water and also through IP injection strongly suggests that increasing the dosage of DOX increases the expression of Cas9 which ultimately impacted on the number of cells mutated. Histopathological analysis suggest the tumors may have a fibroblastic origin owing to the presence of cross striations in herring bone pattern that resembles other previously published fibrosarcoma histology[73]. Additionally, the established sarcoma lines from the tumor tissues were characterized by a fibroblast like morphology suggesting that fibroblasts from the mesenchymal tissues may be the cell of origin for the tumor development. Sequencing the DNA from the established cell lines for mutations in the region surrounding the guide RNA target site in *Trp53*, *Lkb1* and *Kras* revealed anomalies in the *Lkb1* gene compared to the same reads from wild type muscle cell line and the reference sequence. Although mutations are most likely to be present in *Trp53* and *Kras*, we were not able to identify significant aberrations in them through normal sanger sequencing. The nature of the tumor potentially indicating fibrosarcoma, suggests a possibility that coordinated loss of function mutations in Trp53 and Lkb1 and gain of oncogenic mutation in Kras might be the driving factor for the observed phenotype.

This initial pilot study illustrates the power of this approach to initiate sporadic tumor development in mice amidst a healthy immune background and without creating any stable conditional or inducible

knock out mouse lines. Moreover, the tumors that formed were accomplished by site-directed injection of viral gRNAs that utilized a single mouse model rather than breeding multiple transgenic mouse lines to turn on specific oncogenes and at the same time inactivating tumor suppressor genes. The genetic analysis of these tumors and cell lines is ongoing. Most importantly, the cohort of animals that were created by injection with a non-integrating adeno associated viral cassette but not induced with DOX did not develop any tumors. This illustrates the tight control of Cas9 expression from the Col1a1 locus that is dependent upon the induction of DOX for its expression. Compared to the previously available Cas9 knock in mouse models for disease modeling[20], this model has an advantage over them because the expression of Cas9 could be controlled by modulating the levels of DOX induction. This feature would potentially reduce off-target effects associated with ubiquitous Cas9 expression[72]. One issue associated with the delivery of guides through AAV9 or with any other AAV viral vector is that the broad range of tropism exhibited by them[75]. As an example, AAV9 has specificity for lungs, central nervous system, heart and liver apart from skeletal muscles[75]. This might lead to disruption of genes at other tissues when Cas9 is expressed. Although localized delivery of the viral particles to target tissue can circumvent these issues in a considerable manner, there is still a chance that other tissues might also be affected if the viral particles enter the blood stream and travel to transduce distant organs. With regards to this particular aspect, use of the traditional Cre/loxP system for limiting Cas9 expression would allow researchers to achieve the tissue specificity that is currently not present with this iCas9 inducible *in vivo* system. However, breeding a R26-STOP-rtTA^{Flox} (Conditional rtTA) mouse[76] with a tissue specific Cre line would allow researchers to generate a mouse model expressing rtTA only in the specific tissue of interest. This mouse expressing rtTA in a specific tissue of interest when further bred with a Col1a1 - TRE Cas9 mouse would result in the development of offspring wherein the desired Cas9 based gene manipulations would be achieved only in the specific tissue of interest. However, the main limitation with this approach is that this is yet again another triple transgenic mouse model which requires multiple breeding every time a specific tissue of interest has to be manipulated. Alternatively, to reduce the breeding burden, tissue specific rtTA expressing mice could be bred with Col1a1-TRE Cas9 mouse wherein the offspring would be expressing Cas9 only in the specific tissue of interest and only with the administration of doxycycline. Some of the examples of tissue specific rtTA include Albumin rtTA[77] (rtTA expressed only in liver) and Atoh1 rtTA[78] (rtTA expressed only in hair cells). However, the main limitation associated with this approach is that there are very limited number of rtTA specific mouse available compared to tissue specific Cre lines. Nevertheless, compared to the existing transgenic mice models, the DOX inducible Cas9 mouse model is definitely a superior asset to mammalian transgenesis toolbox especially for sporadic tumor modelling.

The established ES cells from this mouse model is an essential and a useful *in vitro* platform for validating the efficacy of different guide RNAs generated that can be used for *in vivo* gene manipulation studies. Although, these iCas9 mES cells express Cas9 only when induced with DOX, the expression levels of Cas9 started to decrease after 24 hrs suggesting the epigenetic silencing of the minimal CMV promoter present within the TRE promoter[26,27]. This factor must be considered when the expression levels of Cas9 needs to be sustained for a long period of time under *in vitro* conditions. The use of T7 Endonuclease assay greatly facilitates the evaluation of gene-edits created by Cas9 at preliminary stages in an efficacious and in an economic manner. However, the low intensity of cleaved bands observed might be because of the varying levels of viral transduction between different cells creating a very heterogenous population of transduced cells. This might be overcome by sorting the transduced cells based upon the BFP expression and using a pure population of lentiviral transduced ES cells for further analysis.

CRISPR/Cas9 based gene modulation of multiple genes has been the basis of the genome wide screening strategies in identifying novel regulators of tumor biology including tumor cell proliferation, metastasis, drug resistance, therapy relapse to name a few[55]. In recent years, the application of CRISPR-Cas9 genome-wide screens *in vivo* PDX based transplantation mouse models and in genetically modified mouse models have gained traction in identifying novel factors pertaining to various aspects of cancer biology. This Cas9 mouse model will be a useful resource for CRISPR-Cas9 based genome wide *in vivo* screens as our sarcoma study has elucidated, albeit on a smaller scale.

One other drawback of this mouse model is that different gene activation/repression fusion proteins must be ectopically delivered into the mouse along with modified guide RNAs for modulating endogenous gene expression in inducible gain/loss of function settings. This may increase or exceed the packaging capacity of a gene delivery vector and may dampen the efficacy of gene delivery.

4.2 Conditional and inducible wt/dCas9-MPH system

In the second Aim of this thesis, we have attempted to generate a cell-specific and inducible expression system for not only Cas9 but also for the MPH transactivation complex to allow simultaneous tumor suppressor inactivation as well as simultaneous oncogene activation. To accomplish this previously published approaches and vectors were used[26,27]. Gateway cloning system has been well developed and characterized in the Haigh lab to merge different elements together in order to create a myriad of plasmid vectors for locus specific targeting in mES cells in an efficacious manner[26,27]. An efficient gene trap compatible docking mES cell line was developed in our lab previously through which the desired gene targeting vectors could be easily targeted into them through a trap coupled recombinase

mediated cassette exchange reaction (RMCE) with nearly 100% precision[26]. The docking mES cell line also denoted as G4-ROSA-antiluc cells contain two heterospecific Frt sites (Frt wt and mutant sites for recombination), antisense luciferase-IRES-Puromycin gene and a promoter less neomycin resistance gene at the R26 locus[79]. When the gene targeting of the incoming vectors happens in the right manner, the antisense luciferase-puromycin containing cassette should be completely removed from the docking site. Lipofectamine based transfection utilizing FlpE recombinase was employed for targeting the conditional and inducible wt/dCas9-MPH vectors into these R26 docking ES cells.

Since both the modified R26 locus in the docking mES cell line and the targeting vector contains two Frt sites and as the FlpE plasmid results in the integration of any sequence between two pairs of Frt sites, the targeting vector gets incorporated into the R26 locus. Here, the selection of the correctly targeted cells happens through restored (neomycin) G418 drug resistance. The resistance is conferred by the proper alignment of the PGK promoter and the ATG start sequence from the COIN wt/dCas9-MPH vector in frame with the neo resistance gene already present in the modified R26 locus of the docking mES cells. This allows to select the correctly targeted cells with almost 100% efficacy. The positive controls i.e., constitutive wtCas9-MPH, inducible wtCas9-MPH and inducible LacZ expressing mES cells that are all valuable *in vitro* controls for analyzing the numerous aspects of the system were all generated by Recombinase Mediated Cassette Exchange reaction (RMCE) based approaches highlighted above.

A large proportion of constitutive wtCas9-MPH cells express EGFP when profiled through FACS strongly suggesting that Cas9-MPH complex should also be expressed in the inducible system as both of these transcripts are expressed from the same TRE promoter and are a part of the same bicistronic cassette. Here, it should be noted that the EGFP is actually a secondary transcript from the TRE promoter which is separated from the Cas9-MPH by an IRES sequence. Previous studies have shown that when IRES sequences are used for separating two different transcripts in a bicistronic vector, the expression of primary transcript is much higher than the expression of the secondary transcript[80]. Thus, the actual levels of Cas9-MPH could potentially be much higher than the levels of EGFP in both the constitutive and DOX inducible systems. In a related experiment, this was corroborated from the higher levels of LacZ expression in comparison to the EGFP expression levels in the DOX treated inducible LacZ cells. This result indicated that a large population of cells expressed LacZ which is the primary transcript from the TRE promoter than the levels of EGFP observed through FACS profiling.

In the inducible wtCas9-MPH expressing cells that are expected to express wtCas9, MPH and EGFP when induced with DOX, around 4% of cells expressed EGFP under both DOX induced and non-

induced conditions. However, in this case only one clone was analyzed and ES cell clones are known to express varying levels of transgene from the TRE promoter[26,27]. This result highlights the need to analyze multiple clones to identify the ones wherein a large population of cells express EGFP and focus on the characterization of these clones further.

The Conditional and inducible (COIN) wt/dCas9-MPH cells when electroporated with a Cre plasmid are expected to lose a loxP site, a STOP cassette and become resistant to puromycin owing to the expression of rtTA-IRES-puromycin resistant gene from the R26 promoter. The non-cre excised control COIN wt/dCas9-MPH cells are expected to be sensitive to puromycin. However, the non cre excised cells were found to be resistant to puromycin even after 12 days of puromycin selection. Moreover, the putative Cre excised COIN wt/dCas9-MPH clones picked after 12 days, had lost the 5' integration band that was confirmed through PCR due to unknown reasons.

4.2.1 The role of puromycin resistance

Several reasons were postulated for the observed continued puromycin resistance in the COIN wt/dCas9-MPH clones after cassette exchange into the R26 locus of the docking mES cells. Initially, it was thought that the original G4 mES cells that were used for creating the docking mES cell line might be resistant to the puromycin drug and may contribute to the observed resistance. However, subjecting these cells to a puromycin kill curve experiment with puromycin concentrations ranging from 1μg/ml to 2μg/ml revealed that these G4 mES cells are indeed sensitive to the drug at 1.25 μg/ml and were completely wiped out by 6th day of selection affirming that they behave as expected.

At this stage, it was confirmed that the cells become resistant to puromycin after targeting the COIN wt/dCas9-MPH vector into the docking mES cells. It was thought that the FlpE plasmid with a puromycin resistant gene used during RMCE-mediated gene targeting of the COIN wt/dCas9-MPH vectors into the R26 locus of the docking mES cells might have integrated into the genome of the mES cells. Although this event is less likely to have happened, this may be a reason for the observed puromycin resistance. In order to confirm this event, a new FlpE plasmid without puromycin resistance gene was used to target the COIN dCas9-MPH gene targeting vector into the docking mES cells. Similar to the previous gene targeting, the targeted cells were picked after 12 days of G418 resistance and they were verified for correct 5' integration of the vector into the docking site. The cells were then subjected to puromycin kill curve with a concentration ranging from 1μg/ml to 2μg/ml and the cells were still resistant to puromycin media. This has illustrated that the puromycin resistance doesn't stem from the FlpE plasmid.

At this stage, we postulated that the puromycin resistance might stem during the process of RMCE based vector targeting into the R26 locus of the docking mES cells. The docking mES cells contain a puromycin resistance cassette in the antisense orientation in the docking site of the modified R26 locus. During the targeting of the COIN wt/dCas9-MPH gene targeting vectors into the docking site of mES cells, the puro cassette that is expected to be removed from the docking site might have reversed its orientation and integrated into another/neighboring location of the targeted locus. This kind of integration of the existing cassette in the docking mES cells into the neighboring location of targeted locus during gene targeting has also been previously discussed by Belteki et al in describing partially recombined clones using the fC31 integrase system[81]. In this case, during puromycin selection, this effect would lead to the enrichment of mES cells that have the puromycin cassette in them which thwarts the ability to pick correctly Cre excised puromycin resistant cells. This strongly suggests the need for Southern blotting to screen out the targeted clones with such mosaic integration events. Alternatively, the use of another Cre plasmid with a hygromycin resistance gene to remove the STOP cassette should be used to select the correctly Cre excised cells with hygromycin resistance.

Despite the fact that a limited number of COIN ES cell clones were analyzed, there did seem to be a difference in the ability to induce Cas9 expression in this system where the rtTA and tet(o)-Cas9 was present on the same allele (*in cis*) and the iCas9 system when the rtTA and tet(o)-Cas9 alleles are on separate chromosomes (*in trans*). This may be one reason for the robust expression of iCas9 when induced with DOX as rtTA (R26 locus) acts *in trans* with the TRE promoter (col1a1 locus). This methodology of separating the two interacting components to different loci (trans acting elements) seems to increase the expression of the genes of interest. The Haigh lab previously evaluated the expression of OSKM (*Oct4*, *Sox-2*, *Klf-4* and *c-Myc*) factors from a reprogrammable mouse that expressed both rtTA and *tet(o)*-OSKM from the same R26 locus[27] (in cis) and compared it with the OSKM expression from another reprogrammable mouse model that expressed the *tet(o)*-OSKM from the Col1a1 locus and rtTA from the R26 locus (in trans)[82]. This experiment demonstrated that the OSKM expressed *in trans* was much higher than the expression obtained *in cis*. This suggests that separating two interacting components (for example rtTA and tet O) in two different locations of different chromosomes might be more beneficial for drug inducible systems.

Chapter 5: Conclusion and future directions

5.1 Conclusions for Cas9 based tumor modeling

The multiplexing ability of Cas9 endonuclease to conjugate with different guide RNAs at the same time enables to create desired mutations in multiple genes simultaneously for complex genetic disease modeling in a more simplistic manner compared to breeding multiple conditional knock outs/transgenics. This feature enables the creation of mouse cohorts within a short period of time and reduces the generation and culling of mice with unwanted alleles. Moreover, the cost of generating these cohorts is also reduced. The use of mouse models wherein Cas9 is embedded in the genome is much more advantageous for modeling diseases as desired *in vivo* gene manipulations could be achieved just with the delivery of single guide RNAs. Our first approach demonstrated for the first time that a mouse model expressing Cas9 in an inducible manner can be utilized for *in vivo* tumorigenesis. In our case, we demonstrated the ability to alter the fate of normal cells to acquire tumorigenic properties in the presence of the intact immune system when the animals are injected with the AAV virus containing guides for targeting *Trp53*, *Lkb1* and *Kras* and when the Cas9 expression is induced with the administration of a drug. Our model was created without any breeding schemes employed and only a very limited number of animals were used. Our study suggests that this drug inducible Cas9 mouse model would be a valuable tool for modeling adult tumorigenesis originating from any tissue. Although not evaluated in this thesis, a similar tumor modelling approach using Cas9 knock in mouse model illustrated that tumor heterogeneity could be replicated as observed in humans. The main caution associated with this type of approach is the lack of tissue specificity over the expression of Cas9 which may cause off target effects. Moreover, the mode of guide RNA delivery must be carefully chosen to avoid delivery vehicle related issues. Viral vectors-based deliveries though efficient in transducing the cells may create occasional side effects such as insertional mutagenesis in case of lentivirus or inflammation related issues in case of adeno virus. Although less efficient, non viral methodology such as liposomes or nanoparticles may be used for targeted delivery of the guides for reducing the off-target effects. *In vivo* Cas9 based gene activation though still at infancy stages has illustrated its ability to recapitulate gene expression levels as seen in some pathology observed in human cases. In lieu of this, our 2nd approach aimed to create a mouse model for simultaneous gene activation and knock outs. Our methodology to utilize ES cell based transgenesis to create a mouse model for simultaneous gene activation and gene knock outs saw some success at the *in vitro* level with the inducible wtCas9-MPH expressing mES cells expressing EGFP under DOX induced and non induced conditions. This approach attempts to reduce the off-target effects associated with the Cas9

by restricting the Cas9 and the activation protein's expression to the desired tissue and be expressed only in a timed manner.

5.2 Future directions

In the first aim, this thesis investigated the gene-editing capability of Cas9 in the presence of doxycycline. Experiments to demonstrate the ability of this system to activate the desired genes by delivering shorter versions of guide RNAs (around 15 bp) also known as dead guide RNAs with MS2 loops and the MPH gene activation complex would be beneficial to utilize this mouse model for endogenous gene activation purposes. Additional clones of the conditional inducible COIN wt/dCas9-MPH must be profiled by cytometric analysis following Cre-mediated expression in the presence of DOX to identity the clones expressing the highest Cas9 and MPH expression levels. Once such clones are identified then parental COIN-MPH-Cas9-IRES-EGFP-Luciferase mice can be generated and validated for sarcoma formation potential in a similar manner as was performed for the iCas9 model.

References

1. Ericsson AC, Crim MJ, Franklin CL. ŽĨ ŶŝŵĂů M ŽĚĞůŝŶŐ. 2013;(June):201-205.

2. Manuscript A, Era G. Life in the Fast Lane: Mammalian Disease Models in the Genomics Era. *Cell*. 2012;148(6):1099-1109. doi:10.1016/j.cell.2012.02.023.Life

3. Rangarajan A, Weinberg RA. Comparative biology of mouse versus human cells: Modelling human cancer in mice. *Nat Rev Cancer*. 2003;3(12):952-959. doi:10.1038/nrc1235

4. Hanahan D, Weinberg RA. Hallmarks of cancer: The next generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013

5. Richmond A, Yingjun S. Mouse xenograft models vs GEM models for human cancer therapeutics. *DMM Dis Model Mech*. 2008;1(2-3):78-82. doi:10.1242/dmm.000976

6. Van Dyke T, Jacks T. Cancer modeling in the modern era: Progress and challenges. *Cell*. 2002;108(2):135-144. doi:10.1016/S0092-8674(02)00621-9

7. Huijbers IJ, Krimpenfort P, Berns A, Jonkers J. Rapid validation of cancer genes in chimeras derived from established genetically engineered mouse models. *BioEssays*. 2011;33(9):701-710. doi:10.1002/bies.201100018

8. Frese KK, Tuveson DA. Maximizing mouse cancer models. *Nat Rev Cancer*. 2007;7(9):645-658. doi:10.1038/nrc2192

9. Mayer et al. 2012. 基因的改变NIH Public Access. *Bone*. 2012;april(12(4)):276-285. doi:10.1038/jid.2014.371

10. Frese KK, Tuveson DA. Maximizing mouse cancer models. *Nat Rev Cancer*. 2007;7(9):645-658. doi:10.1038/nrc2192

11. Kersten K, Visser KE, Miltenburg MH, Jonkers J. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol Med*. 2017;9(2):137-153. doi:10.15252/emmm.201606857

12. Hanahan D, Wagner EF, Palmiter RD. The origins of oncomice: A history of the first transgenic mice genetically engineered to develop cancer. *Genes Dev*. 2007;21(18):2258-2270. doi:10.1101/gad.1583307

13. Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P. Disruption of

overlapping transcripts in the ROSA βgeo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci U S A*. 1997;94(8):3789-3794. doi:10.1073/pnas.94.8.3789

14. Ledermann B. Gene Manipulation and Integrative Physiology Embryonic stem cells and gene targeting. *Exp Physiol*. 2000;85:603-613.

15. Khaled WT, Liu P. Seminars in Cell & Developmental Biology Cancer mouse models : Past , present and future. 2014;27:54-60.

16. Bouabe H, Okkenhaug K. Europe PMC Funders Group Gene Targeting in Mice : a Review. *J Innate Immun*. 201AD;1064(September):315-336. doi:10.1007/978-1-62703-601-6

17. Cheon DJ, Orsulic S. Mouse models of cancer. *Annu Rev Pathol Mech Dis*. 2011;6:95-119. doi:10.1146/annurev.pathol.3.121806.154244

18. Reece-Hoyes JS, Walhout AJM. Gateway recombinational cloning. *Cold Spring Harb Protoc*. 2018;2018(1):1-6. doi:10.1101/pdb.top094912

19. Turan S, Galla M, Ernst E, et al. Recombinase-mediated cassette exchange (RMCE): Traditional concepts and current challenges. *J Mol Biol*. 2011;407(2):193-221. doi:10.1016/j.jmb.2011.01.004

20. Platt RJ, Chen S, Zhou Y, et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell*. 2014;159(2):440-455. doi:10.1016/j.cell.2014.09.014

21. Kim H, Kim M, Im S-K, Fang S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res*. 2018;34(4):147. doi:10.5625/lar.2018.34.4.147

22. T. Das A, Tenenbaum L, Berkhout B. Tet-On Systems For Doxycycline-inducible Gene Expression. *Curr Gene Ther*. 2016;16(3):156-167. doi:10.2174/1566523216666160524144041

23. Friedrich G, Soriano P. Promoter traps in embryonic stem cells: A genetic screen to identify and mutate developmental genes in mice. *Genes Dev*. 1991;5(9):1513-1523. doi:10.1101/gad.5.9.1513

24. Turan S, Zehe C, Kuehle J, Qiao J, Bode J. Recombinase-mediated cassette exchange (RMCE) - A rapidly-expanding toolbox for targeted genomic modifications. *Gene*.

2013;515(1):1-27. doi:10.1016/j.gene.2012.11.016

25. Nyabi O, Naessens M, Haigh K, et al. Efficient mouse transgenesis using Gateway-compatible ROSA26 locus targeting vectors and F1 hybrid ES cells. *Nucleic Acids Res*. 2009;37(7). doi:10.1093/nar/gkp112

26. Haenebalcke L, Goossens S, Naessens M, et al. Efficient ROSA26-based conditional and/or inducible transgenesis using RMCE-compatible F1 hybrid mouse embryonic stem cells. *Stem Cell Rev Reports*. 2013;9(6):774-785. doi:10.1007/s12015-013-9458-z

27. Haenebalcke L, Goossens S, Dierickx P, et al. The ROSA26-iPSC Mouse: A Conditional, Inducible, and Exchangeable Resource for Studying Cellular (De)Differentiation. *Cell Rep*. 2013;3(2):335-341. doi:10.1016/j.celrep.2013.01.016

28. Takahashi K, Yamanaka S. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*. 2006;126(4):663-676. doi:10.1016/j.cell.2006.07.024

29. Renaud-Gabardos E. Internal ribosome entry site-based vectors for combined gene therapy. *World J Exp Med*. 2015;5(1):11. doi:10.5493/wjem.v5.i1.11

30. Mali P, Esvelt KM, Church GM. Cas9 for engineering biology. 2013;10(10):957-963. doi:10.1038/nmeth.2649.Cas9

31. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-)*. 2012;337(6096):816-821. doi:10.1126/science.1225829

32. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*. 2013;8(11):2281-2308. doi:10.1038/nprot.2013.143

33. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-)*. 2014;346(6213). doi:10.1126/science.1258096

34. La Russa MF, Qi LS. The New State of the Art: Cas9 for Gene Activation and Repression. *Mol Cell Biol*. 2015;35(22):3800-3809. doi:10.1128/mcb.00512-15

35. Gräslund T, Li X, Magnenat L, Popkov M, Barbas CF. Exploring strategies for the design of artificial transcription factors: Targeting sites proximal to known regulatory regions for the induction of γ-globin expression and the treatment of sickle cell disease. *J Biol Chem*.

2005;280(5):3707-3714. doi:10.1074/jbc.M406809200

36. Emami KH, Carey M. A synergistic increase in potency of a multimerized VP16 transcriptional activation domain. *EMBO J*. 1992;11(13):5005-5012. doi:10.1002/j.1460-2075.1992.tb05607.x

37. Kimura H. Histone modifications for human epigenome analysis. *J Hum Genet*. 2013;58(7):439-445. doi:10.1038/jhg.2013.66

38. Thakore PI, Black JB, Hilton IB, Gersbach CA. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods*. 2016;13(2):127-137. doi:10.1038/nmeth.3733

39. Schmitz ML, Baeuerle PA. The p65 subunit is responsible for the strong transcription activating potential of NF-χB. *EMBO J*. 1991;10(12):3805-3817. doi:10.1002/j.1460-2075.1991.tb04950.x

40. Staudt MR, Dittmer DP. The Rta/Orf50 Transactivator Proteins of the Gamma-Herpesviridae. In: Boshoff C, Weiss RA, eds. *Kaposi Sarcoma Herpesvirus: New Perspectives*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007:71-100. doi:10.1007/978-3-540-34344-8_3

41. Chavez A, Scheiman J, Vora S, et al. Highly efficient Cas9-mediated transcriptional programming. *Nat Methods*. 2015;12(4):326-328. doi:10.1038/nmeth.3312

42. Silvana Konermann1, 2, 3, 4,*, Mark D. Brigham1, 2, 3, 4,*, Alexandro E. Trevino1, 2, 3, 4, Julia Joung1, 4, Omar O. Abudayyeh1, 2, 3, 4, Clea Barcena1, 2, 3, 4, Patrick D. Hsu1, 2, 3, 4, Naomi Habib1, Jonathan S. Gootenberg1, 2, 3, 4, 5, Hiroshi Nishimas and F, Zhang1, Konermann S, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. 2015;517(7536):583-588. doi:10.1038/nature14136.Genome-scale

43. Zhou H, Liu J, Zhou C, et al. In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR-dCas9-activator transgenic mice. *Nat Neurosci*. 2018;21(3):440-446. doi:10.1038/s41593-017-0060-6

44. McCarty NS, Graham AE, Studená L, Ledesma-Amaro R. Multiplexed CRISPR technologies for gene editing and transcriptional regulation. *Nat Commun*. 2020;11(1):1-13. doi:10.1038/s41467-020-15053-x

45. Cheng AW, Wang H, Yang H, et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res*. 2013;23(10):1163-1171.

doi:10.1038/cr.2013.122

46. Matjusaitis M, Wagstaff LJ, Martella A, et al. Reprogramming of Fibroblasts to Oligodendrocyte Progenitor-like Cells Using CRISPR/Cas9-Based Synthetic Transcription Factors. *Stem cell reports*. 2019;13(6):1053-1067. doi:10.1016/j.stemcr.2019.10.010

47. Zalatan JG, Lee ME, Almeida R, et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell*. 2015;160(1-2):339-350. doi:10.1016/j.cell.2014.11.052

48. Urrutia R. KRAB-containing zinc-finger repressor proteins. *Genome Biol*. 2003;4(10). doi:10.1186/gb-2003-4-10-231

49. Dahlman JE, Abudayyeh OO, Joung J, Gootenberg JS, Zhang F, Konermann S. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol*. 2015;33(11):1159-1161. doi:10.1038/nbt.3390

50. Kiani S, Chavez A, Tuttle M, et al. Cas9 gRNA engineering for genome editing, activation and repression. *Nat Methods*. 2015;12(11):1051-1054. doi:10.1038/nmeth.3580

51. Chakraborty S, Ji H, Kabadi AM, Gersbach CA, Christoforou N, Leong KW. A CRISPR/Cas9-based system for reprogramming cell lineage specification. *Stem Cell Reports*. 2014;3(6):940-947. doi:10.1016/j.stemcr.2014.09.013

52. Black JB, Adler AF, Wang HG, et al. Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. *Cell Stem Cell*. 2016;19(3):406-414. doi:10.1016/j.stem.2016.07.001

53. Truong VA, Hsu MN, Kieu Nguyen NT, et al. CRISPRai for simultaneous gene activation and inhibition to promote stem cell chondrogenesis and calvarial bone regeneration. *Nucleic Acids Res*. 2019;47(13). doi:10.1093/nar/gkz267

54. Hsu MN, Huang KL, Yu FJ, et al. Coactivation of Endogenous Wnt10b and Foxc2 by CRISPR Activation Enhances BMSC Osteogenesis and Promotes Calvarial Bone Regeneration. *Mol Ther*. 2020;28(2):441-451. doi:10.1016/j.ymthe.2019.11.029

55. He C, Han S, Chang Y, et al. CRISPR screen in cancer: status quo and future perspectives. *Am J Cancer Res*. 2021;11(4):1031-1050. http://www.ncbi.nlm.nih.gov/pubmed/33948344%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC8085856.

56. Loganathan SK, Schleicher K, Malik A, et al. Rare driver mutations in head and neck squamous cell carcinomas converge on NOTCH signaling. *Science*. 2020;367(6483):1264-1269. doi:10.1126/science.aax0902

57. Mou H, Kennedy Z, Anderson DG, Yin H, Xue W. Precision cancer mouse models through genome editing with CRISPR-Cas9. *Genome Med*. 2015;7(1):1-11. doi:10.1186/s13073-015-0178-7

58. Yang H, Wang H, Shivalila CS, Cheng AW, Shi L, Jaenisch R. XOne-step generation of mice carrying reporter and conditional alleles by CRISPR/cas-mediated genome engineering. *Cell*. 2013;154(6):1370. doi:10.1016/j.cell.2013.08.022

59. Mou H, Kennedy Z, Anderson DG, Yin H, Xue W. Precision cancer mouse models through genome editing with CRISPR-Cas9. *Genome Med*. 2015;7(1):1-11. doi:10.1186/s13073-015-0178-7

60. Weber J, Rad R. Engineering CRISPR mouse models of cancer. *Curr Opin Genet Dev*. 2019;54:88-96. doi:10.1016/j.gde.2019.04.001

61. Liao HK, Hatanaka F, Araoka T, et al. In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell*. 2017;171(7):1495-1507.e15. doi:10.1016/j.cell.2017.10.025

62. Liao HK, Hatanaka F, Araoka T, et al. In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell*. 2017;171(7):1495-1507.e15. doi:10.1016/j.cell.2017.10.025

63. Hunt C, Hartford SA, White D, et al. Tissue-specific activation of gene expression by the Synergistic Activation Mediator (SAM) CRISPRa system in mice. *Nat Commun*. 2021;12(1):1-14. doi:10.1038/s41467-021-22932-4

64. Pagliarini R, Shao W, Sellers WR. Oncogene addiction: pathways of therapeutic response, resistance, and road maps toward a cure. *EMBO Rep*. 2015;16(3):280-296. doi:10.15252/embr.201439949

65. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009;6(5):343-345. doi:10.1038/nmeth.1318

66. Lee KH, Chuang CK, Guo SF, Tu CF. Simple and efficient derivation of mouse embryonic

stem cell lines using differentiation inhibitors or proliferation stimulators. *Stem Cells Dev*. 2012;21(3):373-383. doi:10.1089/scd.2011.0021

67. Gorman CM, Merlino GT, Willingham MC, Pastan I, Howard BH. The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eukaryotic cells by DNA-mediated transfection. *Proc Natl Acad Sci U S A*. 1982;79(22 I):6777-6781. doi:10.1073/pnas.79.22.6777

68. Durand S, Cimarelli A. The Inside out of Lentiviral Vectors. *Viruses*. 2011;3(2):132-159. doi:10.3390/v3020132

69. Liu Z, Chen O, Wall JBJ, et al. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci Rep*. 2017;7(1):1-9. doi:10.1038/s41598-017-02460-2

70. Klein R, Ruttkowski B, Knapp E, Salmons B, Günzburg WH, Hohenadl C. WPRE-mediated enhancement of gene expression is promoter and cell line specific. *Gene*. 2006;372(1-2):153-161. doi:10.1016/j.gene.2005.12.018

71. Bowling S, Sritharan D, Osorio FG, et al. An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell*. 2020;181(6):1410-1422.e27. doi:10.1016/j.cell.2020.04.048

72. Zhang XH, Tee LY, Wang XG, Huang QS, Yang SH. Off-target effects in CRISPR/Cas9-mediated genome engineering. *Mol Ther - Nucleic Acids*. 2015;4(11):e264. doi:10.1038/mtna.2015.37

73. Augsburger D, Nelson PJ, Kalinski T, et al. Current diagnostics and treatment of fibrosarcoma -perspectives for future therapeutic targets and strategies. *Oncotarget*. 2017;8(61):104638-104653. doi:10.18632/oncotarget.20136

74. Post SM. Mouse models of sarcomas : critical tools in our understanding of the pathobiology. 2012:1-9.

75. Wu Z, Asokan A, Samulski RJ. Adeno-associated Virus Serotypes: Vector Toolkit for Human Gene Therapy. *Mol Ther*. 2006;14(3):316-327. doi:10.1016/j.ymthe.2006.05.009

76. Belteki G, Haigh J, Kabacs N, et al. Conditional and inducible transgene expression in mice through the combinatorial use of Cre-mediated recombination and tetracycline induction. *Nucleic Acids Res*. 2005;33(5):1-10. doi:10.1093/nar/gni051

77. Li Y, Meng R, Chen X, et al. Generation of H11-albumin-rtTA Transgenic Mice : A Tool for Inducible Gene Expression in the Liver. 2019;9(February):591-599. doi:10.1534/g3.118.200963

78. Zuo J. mice : a tool for inducible gene expression in hair cells of the inner ear. 2014:1-9. doi:10.1038/srep06885

79. Haenebalcke L, Goossens S, Naessens M, et al. Efficient ROSA26-based conditional and/or inducible transgenesis using RMCE-compatible F1 hybrid mouse embryonic stem cells. *Stem Cell Rev Reports*. 2013;9(6):774-785. doi:10.1007/s12015-013-9458-z

80. Mizuguchi H, Xu Z, Ishii-Watabe A, Uchida E, Hayakawa T. IRES-Dependent Second Gene Expression Is Significantly Lower Than Cap-Dependent First Gene Expression in a Bicistronic Vector. *Mol Ther*. 2000;1(4):376-382. doi:10.1006/mthe.2000.0050

81. Belteki G, Gertsenstein M, Ow DW, Nagy A. Site-specific cassette exchange and germline transmission with mouse ES cells expressing φC31 integrase. *Nat Biotechnol*. 2003;21(3):321-324. doi:10.1038/nbt787

82. Carey BW, Markoulaki S, Beard C, Hanna J, Jaenisch R. Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nat Methods*. 2010;7(1):56-59. doi:10.1038/NMETH.1410