The Texas Medical Center Library
DigitalCommons@TMC

Dissertations and Theses (Open Access)

MD Anderson UTHealth Houston Graduate School

8-2020

Artificial Intron Technology To Generate Conditional Knock-Out Mice

Amber N. Thomas-Gordon

Follow this and additional works at: https://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the Biochemistry Commons, Cell Biology Commons, Developmental Biology Commons, Genetics Commons, Genomics Commons, Molecular Biology Commons, Molecular Genetics Commons, Other Cell and Developmental Biology Commons, and the Other Genetics and Genomics Commons

Recommended Citation

Thomas-Gordon, Amber N., "Artificial Intron Technology To Generate Conditional Knock-Out Mice" (2020). *Dissertations and Theses (Open Access)*. 1039.
https://digitalcommons.library.tmc.edu/utgsbs_dissertations/1039

This Thesis (MS) is brought to you for free and open access by the MD Anderson UTHealth Houston Graduate School at DigitalCommons@TMC. It has been accepted for inclusion in Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@TMC. For more information, please contact digcommons@library.tmc.edu.

ARTIFICIAL INTRON TECHNOLOGY FOR GENERATING CONDITIONAL KNOCK-OUT MICE

by

Amber N. Thomas-Gordon, B.S., MBA

APPROVED:

Janice Parker-Thornburg
_____
Janice Parker-Thornburg, Ph.D.
Advisory Professor

Richard Behringer
_____
Richard Behringer, Ph.D.

Carrie Cameron
_____
Carrie Cameron, Ph.D.

Jason Heaney
_____
Jason Heaney, Ph.D.

Vicki Huff
_____
Vicki Huff, Ph.D.

APPROVED:

_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

ARTIFICIAL INTRON TECHNOLOGIES TO GENERATE CONDITIONAL KNOCK-OUT MICE

A

THESIS

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences in

Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

by

Amber N. Thomas-Gordon, B.S., MBA

Houston, Texas

August 2020

DEDICATION

To my son, Jayce Gordon, who has inspired and energized me during the most difficult points of my journey. I want to thank him for his generosity and patience despite how challenging it sometimes seem. His love is unconditional and for that I will be forever grateful. To my mother, Julie Moore, who continues to motivate me by her strength and support for me through all that I have faced. To my grandfather, Archie Thomas Sr., and my grandmother Annie Lee Thomas, who showed me how to anchor myself in truth and to persevere regardless of the giants ahead. To my sister, Taylor Moore, and my brother, Calvin Moore, Jr., for lending a helping hand anytime I have needed them and always making me smile. To my cousin, Raven Thomas, for encouraging me by her resilience and understanding during stressful times.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Jan Parker-Thornburg for your support and patience throughout my stay in your laboratory. I appreciate your willingness to train me although I had minimal scientific benchwork experience. I would also like to thank you for reassuring me that I could do it, as it has refined my confidence in pursuing my goals. To the members of my Advisory Committee, Drs. Richard Behringer, Carrie Cameron, Jason Heaney, and Vicki Huff, thank you for all of the feedback and insight you have provided me and helping to mold this project. It has been a privilege to work with you and thank you for serving as a member of my committee.

I am also thankful for my lab mates, Charlie Luo, Hakam Jiroudy, Jennifer Alana, Chad Smith and last but not least, Earnessa Edison. Thank you all for answering all of my questions and teaching me how to navigate the bench. I especially want to thank Charlie Luo for taking so much time to help me through my experiments with tissue culture.

Finally, I would like to extend my thanks and gratitude to Jenny Deng for being patient with me and guiding me through designing my experiments and sharing effective techniques about molecular cloning and scientific knowledge in general. I value the encouragement that you provided to me in pursuing my goals.

ARTIFICIAL INTRON TECHONOLOGY TO GENERATE CONDITIONAL KNOCK-OUT MICE

Amber N. Thomas-Gordon, B.S., MBA

Advisory Professor: Janice Parker-Thornburg, Ph.D.

Genetic engineering has been re-shaped by the invention of new tools in modern biotechnology in a way that offers precision and efficiency in modifying the genome at a single nucleotide level and/or allowing precise control of gene expression. Such gene manipulation brings about significant findings and revelations in comprehending more about embryonic development, cellular and physiological functions, and disease pathology. Current methods used to produce conditional knockouts have limitations on conditional allele placement and modification varies among genes in different organisms. Thus, a system for generating conditional alleles with fidelity remains a challenge. My goal was to examine an approach for generating conditional inactivation of targeted genes that uses a small standardized intron that should be easily introduced using CRISPR/Cas9 methodology. I tested the DECAI (**DE**gradation based on **C**re-regulated-**A**rtificial **I**ntron) artificial intron technology, designed by Dr. Tilmann Burckstummer of Horizon Discovery Ltd., in Vienna Austria, in mouse embryonic stem cells (mESC). Given similarities of the splicing machinery between human and mouse, I expected that the artificial intron components would successfully produce conditional alleles of mouse genes. This intron is designed with l*oxP* sites flanking the branch point sequence, removal of which will abrogate splicing and expose translational stop sequences. This small cassette was inserted into the 5' portion of the coding region of the Neomycin

resistance gene (Neo^R) and this construct was randomly integrated into mouse ES cells. To test its utility for gene inactivation, a Hygromycin resistance plasmid containing Cre-recombinase was introduced into the intron-containing mESCs. A plus/minus screen for Neomycin resistance was performed and Sanger based sequencing was done to identify whether a conditional knock-out had been generated. My results show that the Neo^R gene and the artificial intron remained intact and conditional gene inactivation did not occur. However, with thorough investigation of the gene of interest and splicing requirements, this artificial intron technology has potential for practical use in various model systems.

TABLE OF CONTENTS

LIST OF FIGURES

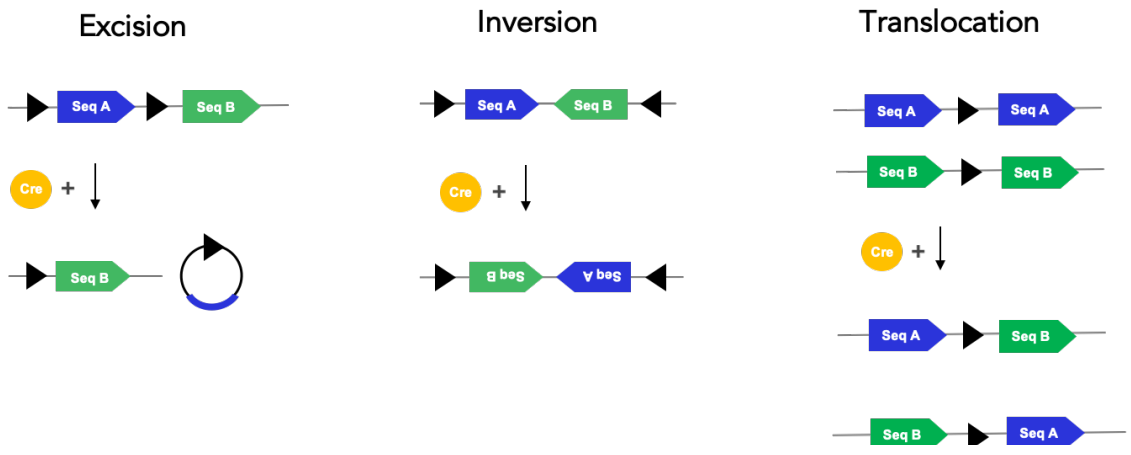
LIST OF TABLES

INTRODUCTION

Genetic engineering has played a vital role in recognizing and understanding what biological processes take place during development and how these processes may translate into disease. It is through genetic modifications that new revelations have emerged for many different organisms. Gene targeting has improved the manner whereby we can determine various biological roles, commonly achieved by producing gene knockouts in the mouse system. Gene knock-out mice have served as a valuable resource for studying embryonic development, typical physiological homeostasis, and human genetic diseases (1).

Conventional gene targeting may identify critical genes within an organism. However, this approach might limit investigators from gaining valuable insight regarding phenotype interpretations because it can lead to lethality mutations (2). Conditional gene knockouts allow for examination of the impact of removing a tissue-specific gene in a spatiotemporal fashion. This method provides better control of where and when the gene will be deleted in comparison to traditional gene knockouts, which may lead to embryonic lethality upon gene inactivation.

Boundaries widened for genome editing technology as the Cre-*loxP* system was discovered as an efficient approach for conditional transgenesis, by allowing for specific gene activation and inactivation (3-4). Cre, a bacteriophage P1-encoded recombinase, is unique in its site-specific recombination system, involving *loxP*, a 34 base pair recognition site which facilitates effectual recombination of the *loxP* sequences, without requiring additional factors (5-6). In this system the gene of interest, flanked by *loxP* sites, may be excised, inverted or translocated, contingent upon the location and orientation of the two *loxP* sites (7), as described in Figure 1 (8).



Used with permission. Image adapted from: Renninger, Sabine & Schonthaler, Helia & Neuhauss, Stephan & Dahm, Ralf. (2011). Investigating the genetics of visual processing, function and behaviour in zebrafish. Neurogenetics. 12. 97-116. 10.1007/s10048-011-0273-x. (8).

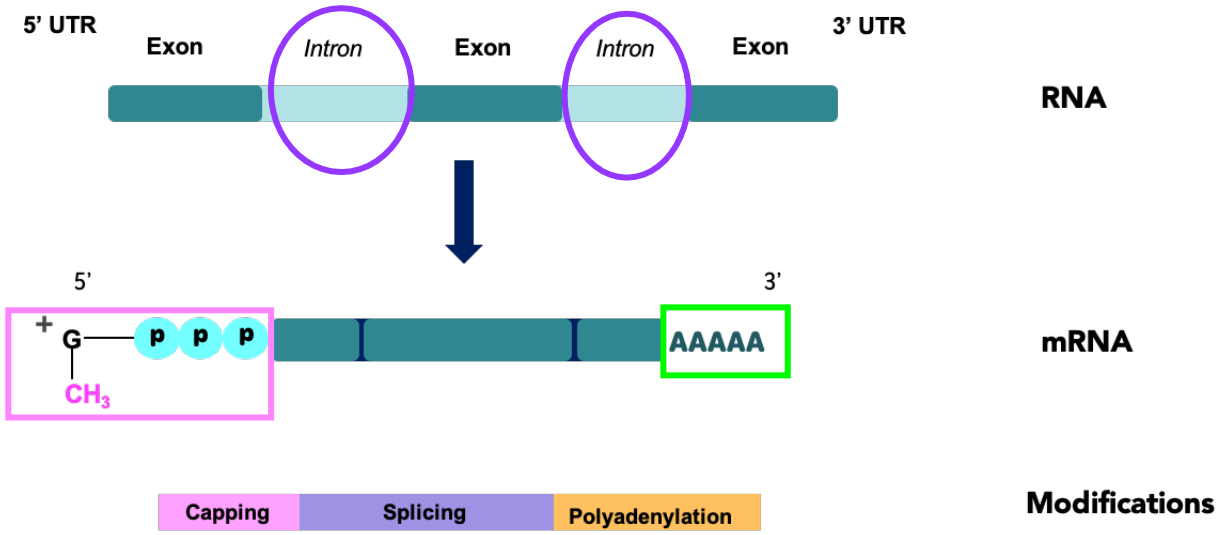
Figure 1. Orientation and location of *loxP* sites produce varying outcomes. Deletions take place when *loxP* sites are oriented in the same direction, leading to DNA excision. Inversions occur when *loxP* sites are located on the same strand of DNA and arranged oppositely resulting in a reversed order of the recombined DNA. . Translocation results from *loxP* sites positioned on separate strands of DNA.

In eukaryotes exons (coding regions) flank non-coding *intervening sequences* (introns), regions of RNA that are spliced out prior to protein translation. Prokaryotic systems maintain a continuous section of coding DNA which is transcribed directly into RNA, as seen in Figure 2 (9). However, the coding DNA in eukaryotes is disjoined into smaller pieces due to the non-coding introns (10). RNA splicing is a post-transcriptional modification event whereby introns are removed from the primary messenger RNA transcript and exons are ligated together to form a mature RNA transcript (11). The mature RNA transcript is then translated into protein.

The traditional method of generating a conditional knock-out using the Cre-*loxP* system involves first flanking the target exon with *loxP* sites oriented in a fashion that

warrants Cre-mediated deletion. Generally, this requires a large construct including *loxP* sites that are oriented for deletion and flanking the critical region in the gene of interest, a selection marker, and homology arms (at a minimum). The construct is introduced into mESCs and selected clones typically require extensive screening for the correct clone(s) prior to introducing it into the whole mouse. The large size of the donor DNA can be difficult and take time to generate and can impact the rate and efficiency of homologous recombination.

To overcome drawbacks in the traditional method aforementioned, artificial introns (AI) can be used as alternative approaches to generate conditional knock-outs. Introns are advantageous because they are capable of being spliced from the gene and allow for the target gene to remain intact. A small intron designed with all critical elements for splicing and that incorporates *loxP* sites flanking critical splicing components within the cassette can potentially improve the rate of insertion and recombination of *loxP* sites. The first reaction to occur in splicing is at the branch point and alteration of the branch point may inhibit the process of splicing. Therefore, flanking the branch point with loxP sites in cis orientation eliminates the branch site sequence and disrupts intron removal. If such an intron cassette is inserted into an exon or coding region of a gene and deletion of the branch point results in exposure to stop sequences, there is potential to inactivate the gene. The coupling of this method with the Cre-*loxP* system may provide a sensible and efficient system to establish a conditional knock-out. Exploring this method in mESCs permits convenient testing of constructs and the ability to screen a large number of cells. In addition, the size of the intron is also small enough to allow for direct in vivo introduction via CRISPR technology.

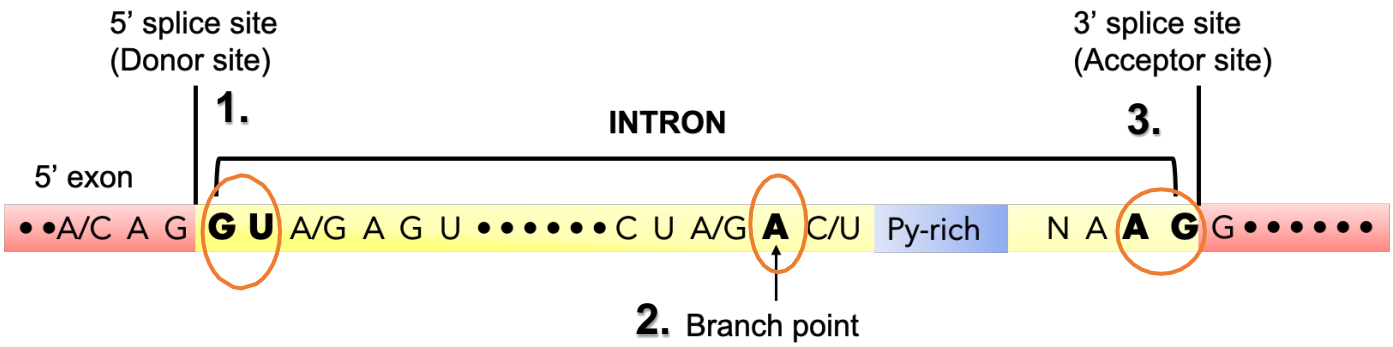


Used with permission. Image adapted from: Alberts B, Johnson A, Lewis J, K. Roberts and P. Walter (eds). Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. From DNA to RNA (9).

Figure 2. Description of post-transcriptional modifications in eukaryotes. The 5' end of the mRNA molecule is capped with a 7-metyl guanosine as indicated by the pink box. The 3' end of the transcript receives a poly-A tail as highlighted by the green box. Splicing removes introns (circled in purple) and exons are then joined together. Once these three post-transcriptional modifications are complete, a newly formed mature RNA is transported to the cytoplasm to be translated into a protein.

There are important aspects surrounding whether splicing successfully occurs and these factors should be especially considered when this process is involved in genetic engineering practices. The complexity of splicing in gene expression is well-organized by the spliceosome, a large multi-protein complex consisting of nearly 200 proteins and five small uridine-rich (U) RNA components—collectively termed snRNPs (12). It is this protein complex that recognizes unique markers at exon-intron boundaries

to activate splicing machinery. Typically, the 5' end (splice donor) of the intron is designated by a GU dinucleotide and the 3' end of the intron contains important conserved sequences such as: the branchpoint (A), a polypyrimidine tract, and the dinucleotide AG (3' splice acceptor) at the most terminal end of the intron (13). Although splice sites and splicing machinery requirements may vary among species, the most common of these are from U2 snRNP and U12 snRNP subsets. The U2 snRNP-dependent group is a major class of introns that are excised by spliceosomes of U1, U2, U4, U5, and U6; Terminal dinucleotides for these introns are *5'*-GT-AG-*3'*, *5'*-GC-AG-*3'*, and *5'*-AT-AC-*3'* (13). See example in Figure 3 below (14).

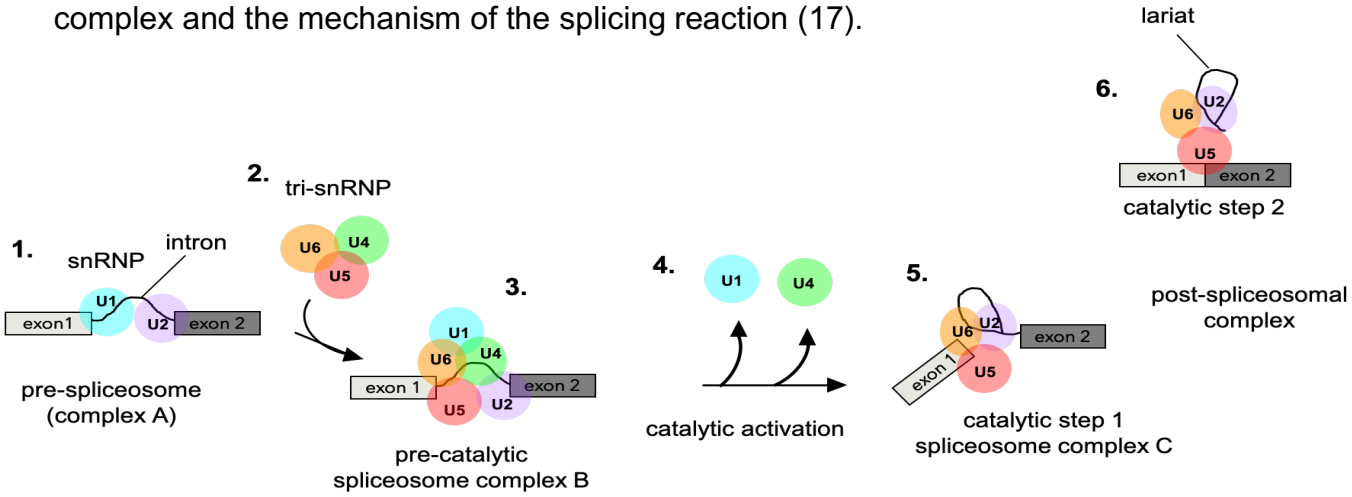


Used with Permission. Image adapted from: Russell, P. (2013). iGenetics: A Molecular Approach. In iGenetics. Pearson Education UK. (14).

Figure 3. Splice site consensus sequences in an intron. The standard RNA molecules are shown. GU (1) marks the 5' splice donor site and (3) AG marks the 3' splice acceptor site. The branch point, A (2), is identified upstream of the polypyrimidine tract (Py-rich).

In addition to splice site markers for initiation of splicing machinery, the branch point sequence (BPS) is also considered a functional aspect of whether particular

biochemical reactions will properly occur to remove the intron. The branch point is located approximately 20-60 nucleotides upstream of the splice acceptor (15) and is a critical part of the initial reaction in splicing. The spliceosome catalyzes this reaction when (i) the 2' hydroxyl group of the branchpoint attacks the first phosphate at the 5' splice donor site forming a branched RNA lariat and (ii) the 3' hydroxyl group of the 5' splice donor site attacks the 3' splice acceptor site resulting in exon ligation and release of the excised RNA lariat (16). Below, Figure 4 shows the assembly of the spliceosome complex and the mechanism of the splicing reaction (17).



Used with permission. Image adapted from: Woll M.G., Naryshkin N.A., Karp G.M. (2017) Drugging Pre-mRNA Splicing. In: Garner A. (eds) RNA Therapeutics. Topics in Medicinal Chemistry, vol 27. Springer, Cham. https://doi.org/10.1007/7355_2017_12. (17).

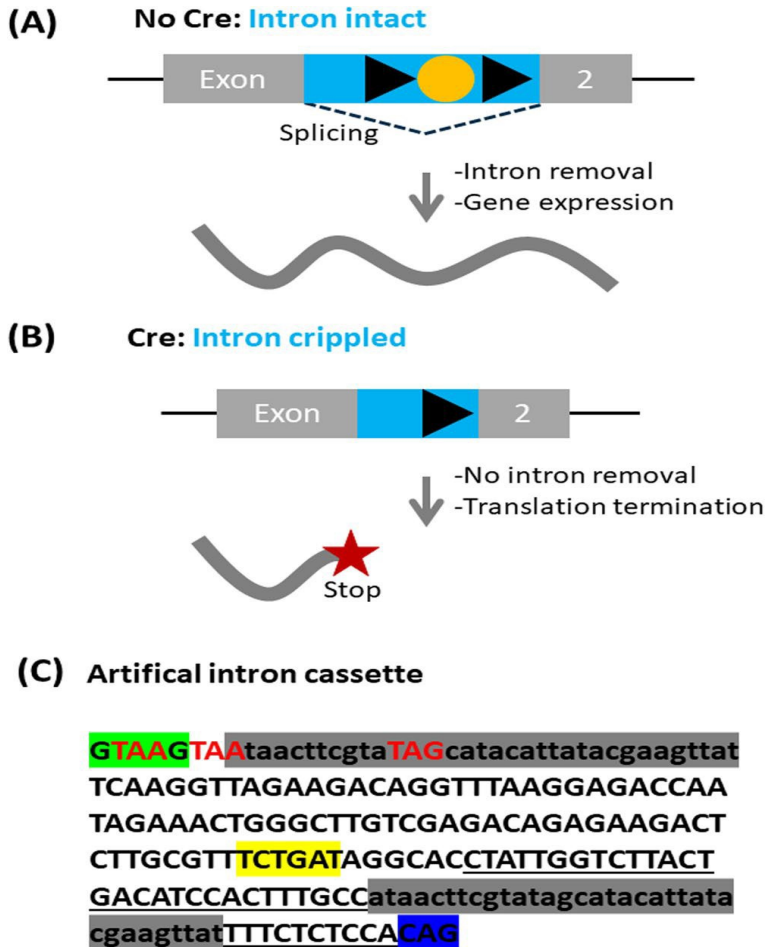
Figure 4. Spliceosome and eukaryotic splicing mechanisms. Small nuclear ribonucleoprotein particles (snRNPs) assemble into a large protein-RNA complex composed of 5 proteins (U1, U2, U4/U6, and U5), known as the spliceosome. Spliceosomes cut out introns and ligate the remaining exons (coding regions) together. Splicing machinery recognizes exon-intron boundaries by a conserved nucleotide sequences in splice sites, preferred intronic sequences and exon enhancers. Assembly of the spliceosome occurs in a step-wise fashion. **(1)** U1 initiates spliceosomal activity by binding to the 5' splice site (donor site) of the intron. Next, U2

is recruited the branch point by other protein factors and binds to the 3' splice site (acceptor site) of the intron. This forms Complex A. A bulge is created in the adenine at the branch point, due to the formation with U2 and the intron. **(2)** Then, tri-snRNP (U6/U4, U5), **(3)** forms with the spliceosome forming Complex B. A structural rearrangement occurs with the addition of U6 and **(4)** ejection of U1 and U4 from the complex. It now becomes catalytically active and **(5)** forming Complex C. Cleavage of the 5' splice site is concurrent with formation of a bond b/t the first nucleotide of the intron and branch point residue. **(6)** The second catalytic step involves the excision of the lariat (intron) and ligation of the 5' and 3' ends of the exons.

Intron excision has been shown to be an important part gene regulation and can impact gene expression. A comparison of gene expression levels tested in both intron-containing cells and intronless cells revealed that intron-containing cells were more transcriptionally efficient ranging from 10-to-100 fold in transgenic mice (18). The influence splicing may have on gene regulation was studied in mice with the H4 promoter linked to a bacterial gene chloramphenicol acetyl transferase (CAT) and it was found that CAT activity was 5-to-300 fold higher in transgenes with the intron compared to those without the intron (19). The Cre-*loxP and FLP-FRT* systems were combined with an intron to develop the COIN (conditionals by inversion) method to generate conditional alleles. The COIN module is inserted into the gene of interest in opposite orientation of transcription direction. No changes are made to the target gene's expression and function until the allele becomes inactivated by either Cre-*loxP* or FLP-FRT. The first module of the COIN allele is an FRT-flanked drug selection cassette that is removed by Flp recombinase. The second module is the functional component consisting of inverted sequences that is flanked by *lox*P sites (head-to-head) and activated by Cre to undergo

inversion. To produce the conditional allele, Cre-recombinase activation causes the COIN module is invert and a reporter gene is expressed and transcription is terminated (20).

In this master's thesis, I focused on investigating the parameters of using the artificial intron technology, DECAI (Degradation based on Cre-regulated Artificial Intron), designed by Tilmann Burckstummer of Horizon Discovery Ltd., (Vienna, Austria), to generate a conditional knock-out in the mouse system. An intron can be inserted into the target gene with all required splicing elements. However, there are multiple stop codons near the 5' end of the intron and upstream to the branch point that is flanked by *loxP* sites. When Cre recombinase is introduced, the branch point is removed and the translational stop sequences are exposed. See Figure 5 for the intron cassette sequence (20). The goal is to introduce a system for targeted gene inactivation that is efficient, reliable, and adaptable across various animal species. The artificial intron cassette under study was introduced into in HAP1 cells and human iPSCs and subsequent to Cre introduction, the sequences between the two loxP sites were removed, as a result of Cre-recombination (20). To assess whether inserting this intron cassette influenced gene expression, cells including the undeleted intron and wild-type were compared. No difference in the level of gene expression was observed. This suggested that the artificial intron did not interfere with endogenous gene expression in the absence of Cre. Upon Cre exposure, both cells (intron-containing and wild-type), showed reduced gene expression relative to the wild-type cells (20).



Used with permission. Guzzardo, P., Rashkova, C., Dos Santos, R., Tehrani, R., Collin, P., & Bürckstümmer, T. (2017). A small cassette enables conditional gene inactivation by CRISPR/Cas9. Scientific Reports, 7(1), 16770–11. https://doi.org/10.1038/s41598-017-16931-z (21).

Figure 5. DECAI approach for conditional gene inactivation. A) The artificial intron (light blue) contains a branch point (yellow circle) flanked by *loxP* sites (black triangles). It is inserted into an exon of a gene. In the absence of Cre, the intron is spliced out and the mRNA remains intact. **B)** In the presence of Cre, the intron is disrupted and will not be spliced out due to the excision of the critical branch point sequence. One of the three translational stop codons in the cassette will cause termination of translation. **C)** Important cassette features: Splice donor (green), *loxP* sites (gray), Branch point

(yellow), Polypyrimidine tract (underlined), Splice acceptor (blue), and 3 translational stop codons (red print).

In this capacity, Cre-recombinase functions as a genetic deactivation switch and therefore, when Cre is not present, splicing machinery removes the intron, translation occurs and a protein product is generated. However, when Cre is present the branch point excised, the intron is not removed, and translation is terminated due to multiple stop codons in the cassette. The goal of my work was to establish the feasibility of using the DECAI artificial intron technology to generate conditional alleles in mouse genes. I also defined whether cell survival was affected due to insertion of the DECAI intron cassette by comparing cell survival of Neomycin resistance in intron-containing and intronless mESCs. Furthermore, I analyzed the capability of the intron to generate a conditional allele in mESCs by introducing a plasmid into the system to disrupt splicing of the intron.

MATERIALS and METHODS

Construction of a plasmid with an artificial intron for insertion into a generic Neomycin resistance gene

A plasmid containing an artificial intron (AI) was designed in house using the technology published by Tillman Burckstummer of Horizon Discovery, LTD, Vienna Austria (20). (Figure 6) The insertion site into the Neomycin resistance gene (Neo^R), bacterial strain, was chosen such that it could subsequently be used in CRISPR experiments. Thus, insertion near a Protospacer adjacent motif (PAM) site was essential. PAM sites were examined through ChopChop (22), a webtool for selecting target sites for the CRISPR system, and selected based on having at least 50% GC content, reduced off-target site, and 60% efficiency. An insertion site for the AI into the Neo^R gene was chosen 110 base pairs from the start site. Figure 6 shows the AI that will be introduced into the Neo^R gene.

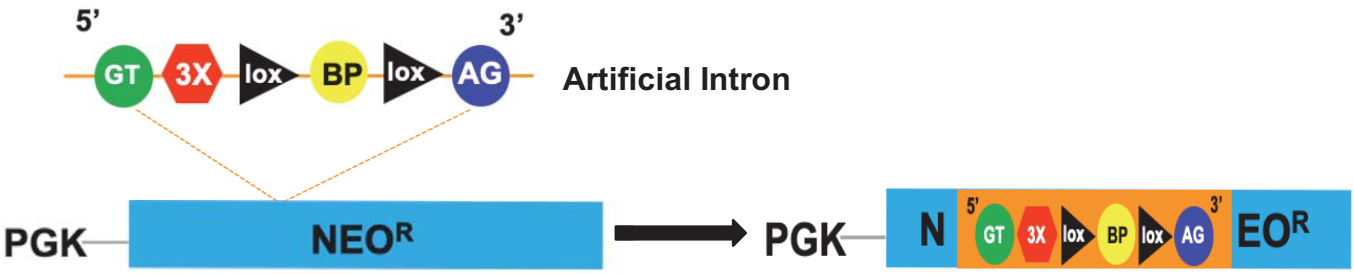


Figure 6. Artificial intron insertion into Neo^R gene. The artificial intron is designed with a splice donor site sequence (GT) at the 5' end of the intron. Three translational stop codons are upstream of the branch point (BP) flanked by two *loxP* sites (lox). The splice acceptor site sequence (AG) is located at the 3' end of the intron. The intron will be inserted into the Neo^R coding sequence to generate PGKNeo^R-AI.

To insert the AI into the gene, two unique restriction enzyme sites, Eag1 and Msc1, were found that spanned the area targeted for insertion. The AI construct was designed to include the same unique restriction enzyme sites for easy insertion into the Neo^R gene. The sequence of the AI (201 base pairs) with flanking sequences of the Neo^R gene, including the Eag1 and Msc1 restriction enzyme sites was sent to Invitrogen by Thermo Fisher Scientific (United States) for commercial synthesis. The total size of the gene synthesized was 434 base pairs (bp).

Isolation of AI fragment and preparation for insertion into Neo^R gene

First, the AI plasmid was submitted for Sanger-based DNA sequencing to the Advanced Technology Genomics Core (ATGC) (MD Anderson Cancer Center, Houston, TX) The sample DNA was submitted at 100 ng/ul, 10 ul per reaction. T3 forward (FW) and reverse (RV) primers were supplied by the ATGC. Qiagen Plasmid Midi-prep Kit protocol (cat. nos. 12143 and 121425) was used for purification and extraction of the Eag1-AI-Msc1 plasmid DNA. To isolate the Eag1-AI-Msc1 DNA fragment, 2 ug of the plasmid was digested in a 50 µl restriction enzyme digestion with both Eag1-HF (New England BioLabs, R3505S) and Msc1 (New England BioLabs, R3505S) restriction enzymes, 10X buffer (New England BioLabs CutSmart), and sterile water at 37℃ overnight. A 1.5% agarose gel identified the expected band size, 434 base pairs, of the DNA fragment. The DNA was precipitated by adding 0.3 volumes of 5M NaCl and 2 volumes of 100% cold ethanol (EtOH) to the DNA sample. Then sample was micro-centrifuged at 13,000 RPM for 10 minutes. The pellet was washed 2X with 70% cold EtOH. The sample was micro-centrifuged at 13,000 RPM for 5 minutes and allowed to dry. This method of DNA precipitation was used in subsequent experiments.

The DNA was resuspended in 30 µl of 0.1 x TE (Tris-EDTA) buffer. Note that preparation of 1X TE buffer with 10mM Tris-HCl, pH 8.0 and 1mM EDTA, pH 8.0 was diluted to 0.1X with sterile water.

Preparation of PGKNeo^{R}bpA and pIBI30 to generate pIBI30-PGKNeo^{R}bpA

A plasmid expressing Neo^R and the PGK promoter (PGKNeo^{R}bpA) was purchased from Addgene and submitted for Sanger based sequencing (to confirm the Neo^R) at 100 ng/µl, 10 µl per reaction. T3 FW and RV primers were supplied by the ATG (Advanced Technology Genomics Core). The PGKNeo^{R}bpA fragment (2 ug) was isolated from the backbone in a 50 µl restriction enzyme digestion with Xho1 (New England BioLabs, R0146S), 10X buffer (New England BioLabs CutSmart), and sterile water at 37°C overnight and the expected band sizes of 1.6 kb and 2.9 kb for the PGKNeo^{R}bpA fragment were identified by 1.5% agarose gel. The fragment was then precipitated by the DNA precipitation method described above and resuspended in 30 ul of 0.1 X TE buffer.

An in-house lab plasmid, pIBI30 (https://www.ncbi.nlm.nih.gov/nuccore/L08878), was used as a cloning vector for the PGKNeo^{R}bpA fragment. A previously unidentified Msc1 site (blunt-end) was confirmed by Sanger based sequencing. Msc1 is a blunt-end site and needed to be removed from the plasmid because it would later interfere with cloning of the AI into Neo^R. Therefore, I located a HindIII (New England BioLabs, R0104S) sticky-end restriction site nearby the Msc1 restriction site in the plasmid and added HindIII to the restriction enzyme digestion. This was done so that Mung bean nuclease would make HindIII a ligatable blunt-end, T4 DNA ligase will be used next to fill in the newly formed blunt-end and destroy the old HindII and Msc1 sites. To correct

the Msc1 site, 1 ug of pIBI30 plasmid DNA was digested in a 20 ul enzyme restriction digestion with Msc1 (New England BioLabs, R3505S), HindIII (New England BioLabs, R0104S), 10X buffer (New England BioLabs CutSmart), and sterile water at 37℃ overnight. A 1.2% agarose gel identified a single band of 2.9 kb. 1 ul of Mung bean nuclease (New England BioLabs M0250S) was added to the digestion above and incubated at 30℃ for 30 minutes. Once the reaction was removed from incubation, DNA was precipitated. To fill-in the plasmid, a 10 ul DNA ligation reaction was performed using 1 ug of DNA, 1 unit (1 ul) of T4 DNA ligase (New England BioLabs M0202S), 10X buffer (New England BioLabs CutSmart), and sterile water. The reaction was incubated for 15 minutes at 12℃. The reaction was removed from incubation and inactivated at 65℃ for 10 minutes. To verify that the Msc1 site had been ablated, a 20ul Msc1 enzyme restriction digest was done using 1 ug of pIBI30-Msc1^{-} Hd3^{-}, 1 ul of Msc1, 10X buffer, and sterile water at 37℃ for 3 hours. At this point, only a single band was detected.

I transformed 75 ul of E. coli DH5-alpha heat shock competent cells with the pIBI30-Msc1^{-}Hd3^{-}. 15 ul of plasmid DNA were mixed with competent cells and incubated on ice for 30 minutes. Once removed from ice, cells were immediately heat shocked for 45 seconds at 42℃ and then placed back on ice for 2 minutes. A solution of 50 ml of LB broth and 50 ul of ampicillin was made, in which 250 ul was added to each tube of cells. Cells were placed in a shaking incubator for 1 hour at 37℃. After cells were removed from incubation, 50 ul and 100 ul of experimental cells and 50 and 100 ul of control cells were plated on pre-warmed (37℃) LB agar ampicillin plates.

Cells were evenly plated and then incubated overnight at 37℃. The following day, plates were removed from incubation and 10 clones were picked. DNA isolation from

these clones was done using the Qiagen mini-prep kit protocol (cat. nos. 12123 and 12125). A 20 ul restriction enzyme digestion was done for each of the 10 clones using HindIII and Xho1 restriction enzymes at 37℃ overnight. A 1.2% agarose gel for clones 3, 8, and 9 verified that Xho1 had a single band size 2.9 kb and HindIII restriction enzyme did not make a cut in the plasmid, indicating that HindIII and Msc1 had been ablated. Clone 9 was chosen to use as the final plasmid backbone for insertion of PGKNeo^{R}bpA.

Xho1 was present in both IBI30 and PGKNeo^{R} making it an ideal restriction site to use for inserting the Neo^{R} into IBI30. To prepare for generating pIBI30-PGKNeo^{R}bpA, an Xho1 restriction enzyme digestion was done for both PGKNeo^{R}bpA and pIBI30-Msc1^{-}Hd3^{-}. Identification of two band sizes of 1.6 kb and 2.9 kb for PGKNeo^{R}bpA and one band size of 2.9 kb pIBI30-Msc1^{-}Hd3^{-} respectively, were confirmed by a 1.5% agarose gel. Each band of was isolated from agarose gel, place in a 15 ml Falcon tube and weighed. Sodium iodide (NaI) was added and the gel melted at 55℃. Then 120 ul of GeneClean beads were added and rotated for 15 minutes and centrifuged at 4.4 x 10^{3} RPM for 10 minutes. The supernatant was removed and the pellet was resuspended in 120 ul of NewWash. The bead solution was split into four filtered tubes and washed twice and then centrifugation for 30 seconds. Filters were transferred tube-to-tube to concentrate the DNA after adding GeneClean Spin Elution buffer (60ul), 5 minutes of 37℃ incubation, and being centrifuged for 30 seconds. Lastly, the DNA was incubated at 37℃ for 5 minutes and centrifuged for 30 seconds. Gel purification yielded 37.8 ng/μl for PGKNeo^{R}bpA and 26 ng/μl of pIBI30-Msc1^{-}Hd3^{-}. To complete the generation of pIBI30-PGKNeo^{R}bpA, a 10 ul ligation reaction was done using T4 DNA ligase with a molar 1:3 ratio of pIBI30-

Msc1^{-}Hd3^{-}vector backbone and PGKNeo^{R}bpA fragment insert. For the ligation, I used 6 ul of PGKNeo^{R}bpA insert, 2 ul of pIBI30-Msc1^{-}Hd3^{-}vector, 1 ul of T4 DNA ligase (New England BioLabs M0202S), 1 ul of 10X buffer (New England BioLabs CutSmart). The reaction was incubated for 15 minutes at 12℃. The reaction was removed from incubation and the ligase inactivated at 65℃ for 10 minutes. Two controls were used each in a 10ul ligation reaction. Control A did not have the insert and control B did not have the insert or T4 DNA ligase.

Introduction of the experimental and control plasmids into E. coli DH5-alpha heat shock competent cells was done for the Experimental: (50 µl and 100µl); Control A: (50/100); Control B (50/100). From transformation, 40 clones were picked and the Qiagen mini-prep kit was used to isolate DNA from 20 clones. A 20 ul restriction enzyme digestion was set up for all 20 clones using Msc1, Xho1, Eag1, and Eag1/Msc1. Expected band sizes for an appropriately ligated plasmid were as follows: Msc1-4.7 kb, Eag1, 4.7 kb, Eag1/Msc1-400pb and 4.3 kb, and Xho1- 1.6 kb and 2.9 b. A diagnostic 1.5% agarose gel confirmed the expected band sizes for only clone #2 of the 20 selected clones. Clone #2 had successfully acquired the PGKNeo^{R}bpA DNA insertion and would be used in subsequent experiments.

Unique restriction enzyme sites, Eag1 and Msc1, allowed for insertion of the AI into the Neo^{R} gene to generate the final plasmid, pIBI30-PGKNeo^{R}bpA-AI. An Eag1 and Msc1 restriction enzyme digestion for pIBI30-PGKNeo^{R}bpA was set up in a 20ul reaction for 2 hrs. at 37° Celsius. A diagnostic gel of 1.2% agarose verified the expected 4.3 kb band size. The band was isolated from agarose gel and gel purification was performed (as described above), yielding 21ng/ul DNA concentration for pIBI30-PGKNeo^{R}bpA.

The AI fragment was isolated from a vector backbone by a 50 ul Eag1 and Msc1 restriction enzyme digestion. The correct band size of 434 bp was identified with a 2% agarose diagnostic gel. The correct band was isolated from agarose gel and gel purification was performed (as described above). A DNA 20 ul ligation reaction was set up (as previously described) with 4 ul of pIBI30-PGKNeo^{R}bpA (100 ng/ul) vector and 3 ul of the AI fragment (60ng/ul), 1 ul of T4 DNA ligase, and 2 ul of 10X buffer. The protocol for bacterial transformation as described above was used for this ligation. Three experimental samples were plated on LB agar ampicillin plates containing 20ul, 50ul, and 100ul of cells and two control plates (insert- and T4-; 50/100 ul per control). A total of 21 colonies grew on the experimental plates and no colonies formed on control plates as expected. DNA was isolated from all 21 colonies via Qiagen plasmid midi-prep kit (cat. nos. 12143 and 121425) protocol and resuspended in 30 ul of 0.1 X TE buffer. 3 ul of DNA was used for each sample in a 20 ul restriction enzyme digestion with Eag1 and Msc1 to check whether restriction sites were intact. This was confirmed by a 1.2% agarose diagnostic gel identifying band sizes of 4.7 kb for both as expected in only clone #7. This newly generated, IBI30-PGKNeo^{R}bpA-AI plasmid was expanded for further analysis in mouse embryonic stem cells (mESC). Sanger based sequencing of the IBI30-PGKNeo^{R}bpA-AI plasmid verified Neo^{R} gene and AI were intact as expected.

Testing artificial intron functionality for Cre deletion in bacterial cells

I transformed 100 ul Cre-expressing bacterial competent cells with 10 ul of the pIBI30-PGKNeo^{R}bpA-AI plasmid (788 ng/ul) to determine the capability of the AI to undergo Cre-recombination. A PGKNeo^{R}-l*oxP*-B (1445.4 ng/ul) plasmid was used as a

control due its success in Cre-recombination in past laboratory projects. Once the plasmids were mixed with the competent cells, they were placed on ice for 30 minutes and immediately heat shocked for 45 seconds at 42℃ and then placed back on ice for 2 minutes. A solution of 50 ml of LB broth and 50 ul of ampicillin was made, in which 250 ul was added to each tube of cells. Cells were placed in a shaking incubator for 1 hour at 37℃. After cells were removed from incubation, 10 ul of experimental cells and 10 ul of control cells were plated on pre-warmed (37℃) LB agar ampicillin plates. Cells were evenly plated and then incubated overnight at 37℃. The following day, plates were removed from incubation and 5 clones were picked for both the experimental and control samples. DNA isolation of these clones were done using the Qiagen mini-prep kit protocol (cat. nos. 12123 and 12125). A 50 ul PCR reaction was conducted for the experimental and control samples Custom forward and reverse primers were used targeting the PGKNeo^{R}bpA-AI fragment. A 1.2% diagnostic gel showed expected band sizes in experimental of 1.6 kb and 1.8 kb in control, due to Cre-recombination.

Compare gene expression levels of Neo^{R} and Neo^{R}-AI in mouse embryonic stem cells (mESC)

To isolate the fragments for introduction into G4 mESC's, 100ug of PGKNeo^{R}bpA-AI was digested in an 800 ul Sal1 restriction enzyme digestion reaction and 100ug of PGKNeo^{R}bpA was digested in an 800 ul EcoR1 reaction. A 1.2% agarose diagnostic gel showed expected band sizes of 4.5 kb for PGKNeo^{R}bpA and 4.7 kb for PGKNeo^{R}bpA-AI. DNA precipitation of each sample yielded a total DNA concentration of 75.2 ug for PGKNeo^{R}bpA and 70 ug for PGKNeo^{R}bpA-AI.

Three trials were conducted to examine how intron insertion would affect gene expression of PGKNeo^{R}bpA and PGKNeo^{R}bpA-AI, individually, in mESCs. 1 x 10^{7} mESCs were electroporated with 25 ug of PGKNeo^{R}bpA and 25 ug of PGKNeo^{R}bpA-AI and cells were plated on 4-10 cm dishes per sample. A positive screen for cells that acquired PGKNeo^{R}bpA and PGKNeo^{R}bpA-AI DNA was conducted with G418 (Gibco Geneticin, ref. 10131-027), 24 hours post-electroporation, was conducted for 10 days and colonies were counted on day 12 for all plates from each trial.

Generation of PGKNeo^{R}bpA-AI mESC line

The plasmid, IBI30-PGKNeo^{R}bpA-AI, was linearized by a 100 ul Sal1 restriction enzyme digestion and a band size of 4.7 kb was verified by a diagnostic gel. Precipitation of the digest yielded a DNA concentration of 546.2 ng/ul of IBI30-PGKNeo^{R}bpA-AI. 25 ug of the plasmid DNA was electroporated into 1 x 10^{7} G4 mESC's and cells were plated on 4-10 cm dishes. Cells were treated with G418 24 hrs. post-electroporation, and grown for 10 days. From this selection, 18 colonies were picked on day 12 and 11 colonies were expanded onto 3-96 well plates. One plate was used for analysis and two back-up plates were frozen down with 50 ul /well of cell freezing media (Bambanker Serum-free Type Cell Freezing Medium, cat no 302-14681) and stored at -80 degrees Celsius. The newly generated IBI30-PGKNeo^{R}bpA-AI mESC line will be used in subsequent experiments for the analysis of the artificial intron functionality.

Generation of PGKNeo^{R}bpA-AI Cre^{+} mESC line

The pMC-Cre PGK-Hygromycin plasmid was linearized by a HindIII (6.5 kb) restriction enzyme digestion at 37℃ overnight. A 1.2% agarose diagnostic gel verified the digestion produced a single cut and the expected band size of 6.5 kb. The DNA was precipitated by (see protocol mentioned above) resulting in a concentration of 340 ng/ul.

25 ug of linearized pMC-Cre PGK-Hygromycin plasmid DNA was electroporated in 1 x 10^{7} PGKNeo^{R}-AI ES cells and plated on 4-10 cm dishes. After 24-hrs of electroporation, a 7-day positive screen for Hygromycin resistance was conducted on the newly generated, PGKNeo^{R}bpA-AI Cre^{+} cells from a single clone. On day nine, 48 clones were picked on a 96 well plate and split to 3-96 well plates, which were stored at -80 ℃. Now, having a total of 4-96 well plates, one plate was used for testing, one for analysis and the two remaining plates were kept for any subsequent experiments.

Further testing of the PGKNeo^{R}bpA-AI Cre^{+} cells for artificial intron functionality was performed by a positive and negative Neomycin resistance screen. The 48 colonies were treated with G418 for 10 days and colonies, both sensitive (Neo^{S}) and resistant (Neo^{R}) to Neomycin selection, were identified and marked accordingly. This selection produced 7 Neo^{S} and 41 Neo^{R} colonies. The analysis plate was thawed and grown until cells reached approximately 70% confluency. Cell colonies were marked as Neo^{S} or Neo^{R} accordingly, on the analysis plate. DNA was extracted directly from the 96-well plate. From the 48 identified clones, 4 out of 7 Neo^{S} clones were chosen for PCR and eventual DNA sequencing, as DNA was lost for 3 Neo^{S} clones during DNA extraction. To be consistent, 4 Neo^{R} clones were randomly selected for PCR and subsequent DNA sequencing.

A PCR reaction (see protocol above) was set up for 4 Neo^S clones, 4 Neo^R clones, PGKNeo^RbpA, and a negative control (water only). A diagnostic gel for 2 Neo^S and 2 Neo^R PCR samples showed a major band size of 401 base pairs for all cell colonies. PCR products for Neo^S and Neo^R clones were submitted for Sanger based DNA sequencing, validating a fully intact Neo^R gene and AI. PCR products were then stored at -20 ℃.

RESULTS

Generation of pIBI30-PGKNeo^{R}bpA-AI

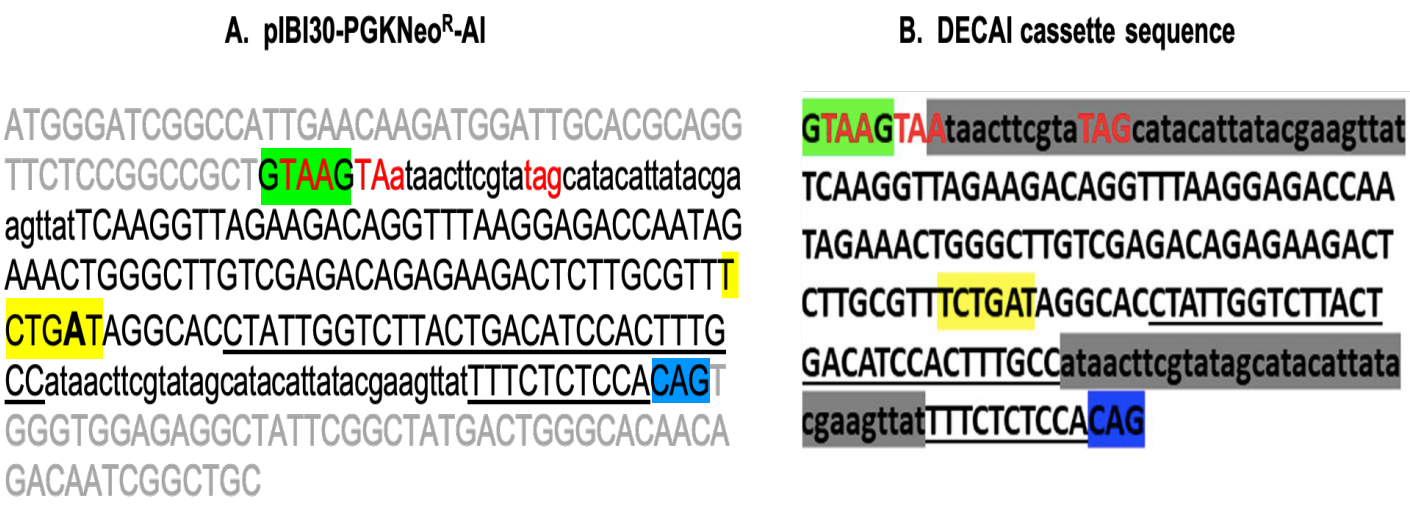
Artificial introns (AI) are, theoretically, a simple, uncomplicated means of generating conditional knockouts. The DECAI intron cassette was tested in HAP1 cells (near-haploid cell line derived from chronic myelogenous leukemia cell line) and human iPSCs (METTL16 essential gene and CD46 non-essential gene) and was shown to be easily transferable between systems/somatic cell lines (21). However, the parameters of use of this artificial intron cassette had yet to be tested in the mouse system.

Prior to use in the intact mouse, this study was performed to verify that the intron would perform in mouse embryonic stem cells (mESCs) in the same manner as it did in human cell lines. To examine this, the Neomycin resistance gene (Neo^R) was used as a test gene because it is simple to screen for the phenotype via drug selection. Neomycin is a part of a class of antibiotics that contains amino sugars joined by glycosidic bonds. Neomycin resistance is conferred by an aminoglycoside phosphotransferase gene. The Neomycin resistance gene is routinely used in mammalian cell culture to express cloned proteins. G418 (as mentioned above) is used in experiments involving eukaryotes to select for neomycin resistance (23).

In order to generate a plasmid with Neo^R containing the artificial intron, I had the intron commercially synthesized from Invitrogen. An in-house lab plasmid, pIBI30, was used as a backbone for the newly generated pIBI30-PGKNeo^R-bpA-AI.

To determine whether the artificial intron was appropriately inserted and that all of the splicing components were unaltered, I first performed Sanger based DNA sequencing to ensure the Neo^R gene and AI cassette were in place as expected. As shown in Figure 7 (21), sequencing demonstrated that the artificial intron was

appropriately inserted and included the intact splice donor, branch point flanked by loxP sites, splice acceptor, and translational stop codons.



Used with permission. Guzzardo, P., Rashkova, C., Dos Santos, R., Tehrani, R., Collin, P., & Bürckstümmer, T. (2017). A small cassette enables conditional gene inactivation by CRISPR/Cas9. Scientific Reports, 7(1), 16770–11. https://doi.org/10.1038/s41598-017-16931-z (21)

Figure 7 Artificial intron sequence of intact cassette. A) Shows that pIBI30-PGKNeo^{R}AI contains the intact intron with all components accounted for. The splice donor is green, the branch point (yellow) is flanked by *loxP* sites (lowercase), the polypyrimidine tract is underlined, the splice acceptor is blue, and three translational stop codons are shown in red. The gray shows the targeted region of the Neo^{R} gene. **B)** Represents the DECAI intron cassette under study. Shown is the splice donor (green), branch point (yellow) flanked by loxP sites (gray), polypyrimidine tract (underlined), splice acceptor (blue), and three translational stop codons(red). Note intronic cassettes for **A** and **B** are equivalent.

I next wished to confirm that the loxP sequences would function as expected to delete the floxed branch point. To test the functionality of the artificial intron, I transformed

Cre-expressing bacteria with the IBI30-PGKNeo^{R}-bpA-AI plasmid. DNA was isolated from the transformed plasmids and compared to the parent plasmid by PCR. I anticipated that the band size of the deleted plasmid would be 1.590 kb and the control plasmid to be 1.737 kb, which was seen upon analysis (Figure 8. This finding indicates that the loxP sites were operative and properly flanked the branchpoint.

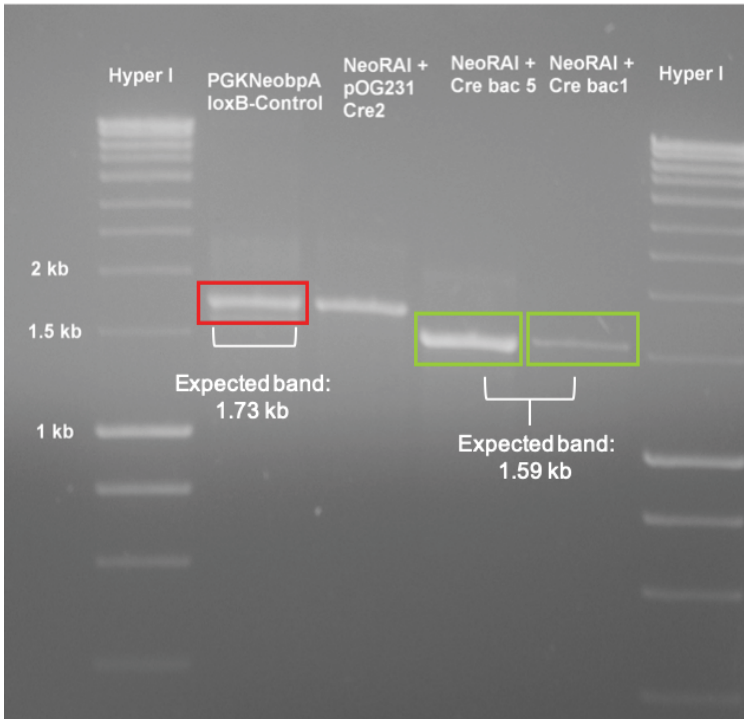


Figure 8. Comparison of Cre-deleted plasmid vs. non Cre-deleted plasmid. pIBI30-PGKNeo^{R}bpA-AI was placed into Cre-expressing bacteria competent cells to test whether *loxP* sites were operable. Neo^{R}AI Cre-bacteria clone 1 and Neo^{R}AI Cre-bacteria clone 5 shows Cre-recombination of the plasmid. Bands for clones 1 and 5 are outlined in green and were expected to be1.59 kb as shown. The PGKNeo^{R}bpA-loxB control plasmid has loxP sites at the end of the gene and the entire NeoR gene will be removed as a result of Cre-recombination. This band is outlined in red and

expected to be 1.73, as shown above. Note that Neo^{R}AI pOG231-Cre 2 was used in error.

Random integration of pIBI30-PGKNeo^{R}bpA-AI into G4 genome of Neo^R ES cell line

Given that the artificial intron was capable of Cre-deletion, the next step was to test the functionality of the intron in mESCs. To prepare the pIBI30-PGKNeo^{R}bpA-AI for insertion into mESCs, I conducted a SalI restriction enzyme digestion to linearize the plasmid. I expected a single band of 4.7 kb and was able to confirm that upon analysis as shown in Figure 9.

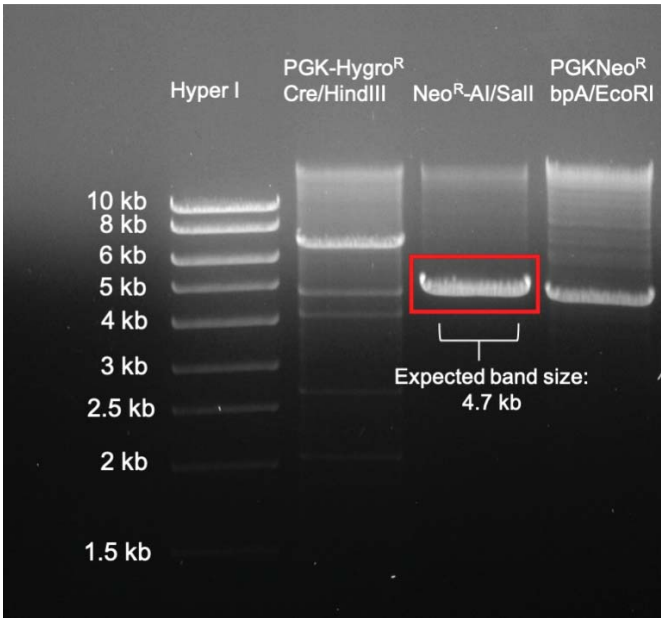


Fig 9. SalI restriction enzyme digestion of pIBI30-PGKNeo^{R}bpA-AI. Plasmid DNA was linearized by Sal1 restriction enzyme prior to electroporation into mESCs (G4). The single band of 4.7 kb for pIBI30-PGKNeo^{R}bpA-AI/Sal1 is outlined in red. Note PGK Hygro^R-Cre plasmid and PGKNeo^{R}bpA were also linearized at this time to test enzyme restriction sites prior to electroporation in mESCs subsequently.

To generate a mouse ES cell line containing the Neo^R-AI plasmid DNA, the DNA was randomly integrated by electroporating 1 x 10^7 G4 cells with the linearized Neo-AI plasmid. As a control, Neo^R was electroplated into 1 x 10^7 cells. A 7-day positive screen for Neomycin resistance was conducted to isolate clones that acquired the DNA for both cohorts (Neo^R and Neo^R-AI). It should be noted that cells containing the Neo^R-AI plasmid had -4.8-fold fewer in clone numbers compared to those of the Neo^R gene without the AI. To quantify the difference, I repeated the experiment, plating both NeoR-AI and NeoR cells onto 4-10 cm plates and counted the surviving colonies. Table 1 presents the total number of cell colonies Neo^R and Neo^R-AI colonies. Although fewer Neo^R-AI colonies

were present, I elected to proceed with experiments and 11 Neo^R-AI colonies were expanded for further testing.

Trial	Number of Neo^R-AI mESC colonies/1 x 10^7 cells electroporated	Number of Neo^R mES cell colonies/1 x 10^7 cells electroporated
1	41	229
2	54	317
3	59	197
	154	743

Table 1. Neo^R mESC colonies vs. Neo^R-AI mESC colonies. Insertion of the AI into the Neo^R gene suggested an impact on gene expression, as visually observed by the number of cell colonies that survived G418 selection. Cells were plated on 4-10 cm dishes for each trial corresponding intron-containing and intronless colonies. Results for from the experiments of the three trials show a -4.8-fold decrease in the number of colonies produced by cells with the artificial intron than those without the artificial intron.

Exposure of Neo^R-AI ES cells to pMC1-Cre, PGK Hygromycin resistance plasmid with Cre expression ($Hygro^R$-Cre)

For further experiments, I randomly selected one of the 11 colonies to expand. The newly generated Neo^R-AI ES cell line was then exposed to Cre recombinase using a Cre-expressing plasmid with hygromycin as a selectable marker. The pMC-Cre, PGK $Hygromycin^R$ Cre ($Hygro^R$-Cre) plasmid DNA was linearized by HindIII restriction enzyme

digestion. It was next purified and electroporated into 1 x 10^7 Neo^R-AI ES cells. I chose to linearize the $Hygro^R$-Cre plasmid to increase the chance of integration into the cells.

To confirm that the $Hygro^R$-Cre plasmid DNA was attained by Neo^R-AI ES cells, a positive screen for hygromycin was conducted for 7 days and 48 hygromycin resistant colonies were selected for further analysis. The $Hygro^R$-Cre plasmid was presumed to have integrated into the cells that survived Hygromycin resistance selection. The surviving clones were expanded onto 4-96 well plates, in which one plate was used for testing, one for analysis, and two for back-ups.

Positive/negative screen on NeoR-AI Cre+ cell colonies for neomycin resistance

I expected that the Neo^R-AI gene, in the presence of the $Hygro^R$-Cre plasmid would be inactivated and Neo^R protein would not be expressed. At this point, DNA sequencing confirmed that the cells initially used contained the Neo^R gene and AI. Based on their resistance to Hygromycin, they should have been exposed to Cre-recombinase. With that exposure, the branch point should be deleted and stop codons exposed. Thus, I expected that these clones would now be sensitive to Neomycin. To test this, cells were treated with G418 for 10 days to test for Neomycin sensitivity. This selection interestingly revealed that 41 colonies survived G148 selection and were neomycin resistant, while only 7 colonies did not survive and were Neomycin sensitive (Neo^S). The DECAI method is centered around branch point removal by Cre-recombination which should disrupt splicing. The translational stop codons in the cassette, upstream of the floxed branch point will be recognized by the translational machinery and terminate translation. These results would suggest that Neo^R-AI gene had not been affected in

spite of presumed exposure to Cre-recombinase, which had been expected to remove the branch point.

Sequence results for Neo^{R}-AI and Neo^{S}-AI cell colonies

Next, I conducted standard PCR on 4 Neo^{R}-AI cell colonies and 4 Neo^{S}-AI cell colonies to confirm appropriate band sizes of 401 base pairs and 254 base pairs, respectively. It was shown that both Neo^{R}-AI and Neo^{S}-AI cell colonies had the same band size of 401 base pairs, suggesting that the AI had not been Cre-deleted and the neomycin resistance gene had not been inactivated in either set of cell colonies, as shown in Figure 10.

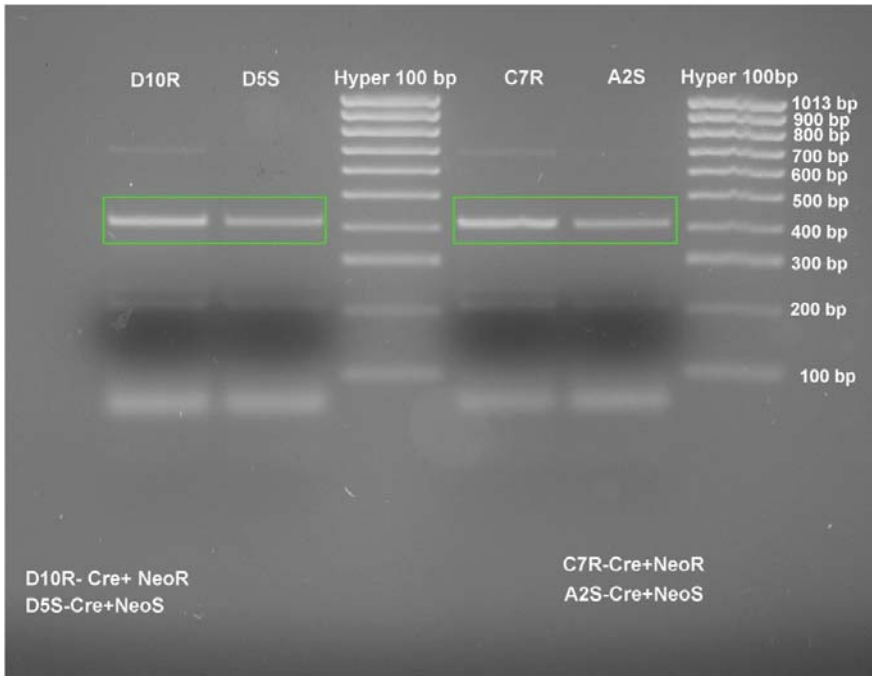


Figure 10. Neo^{R}-AI Cre^{+} cell colonies vs. Neo^{S}-AI Cre^{+} cell colonies PCR. This figure shows four total clones that contain the Neo^{R} gene, artificial intron (AI) and Hygro^{R} Cre plasmid. After Cre integration and the Neomycin resistance screen, cell colonies that did not survive Neomycin resistance (Neo^{S}) selection were expected to be Cre-recombined with a band size of 254 base pairs. Cell colonies that did survive Neomycin resistance

(Neo^R) selection and were not Cre-recombined, and a band size of 401 bp was expected. Band sizes outlined in green show: 2 Neo^S clones (A2S and D5S) and 2 Neo^R clones (C7R and D10R) all with the same band size of 401 base pairs. This indicated that none of these clones were Cre-recombined despite their response to drug selection.

To verify whether the AI and neomycin resistance gene were fully intact, Sanger based DNA sequencing was done. Sequencing revealed that Neo^R-AI and Neo^S-AI colonies that came through this screen carried a non-mutated AI and Neo^R gene (Figure 11). From these results, it was clear that Neo^S-AI cell colonies had not been Cre-recombined and branchpoint deletion was not detected.

Figure 11. DNA sequencing results of Neo^R-AI Cre+ cell colonies and Neo^S-AI Cre+ cell colonies. Cell colonies containing the intron and that were sensitive and resistant to Neomycin upon drug selection, both contained an intact intron and retained the Neo^R

gene 100%. All intronic components are consistent with after having been introduced to Cre-recombinase: splice donor (green), flanked branch point (yellow) by loxP sites (lowercase), polypyrimidine tract (underlined), splice acceptor (blue), and translational stop codons (red). Gray area shows the Neo^R gene that was targeted. The DNA sequencing verified that there were no Cre-recombined cells.

DISCUSSION AND CONCLUSION

The most ideal strategy for regulating conditional gene expression would include a genetic switch that controls fundamental and increased levels of gene expression, is reversible and does not alter other cellular functions (24). Although we have not reached such a method, modern approaches using artificial intron technology are increasingly utilized and proving to be effective in conditional gene expression. The DECAI method used in HAP1 cells and human iPSCs demonstrated significant decrease in gene expression (21) and shows potential for application to various species.

The intron cassette under study is advantageous in that (i) it is small in size (201 bp) (ii) the simple design includes the most critical components of an intron and that are necessary for the splicing reaction (iii) it is constructed with multiple translational stop codons to ensure nullification of gene expression (iv) it uses the Cre-*loxP* system which is straightforward and does not require additional factors to function (25) (v) it's capable of splicing out as an endogenous intron without altering cellular functions and (vi) it may be simply delivered into the genome by CRISPR technology. Based on its seemingly convenient use, I believed the technology could be optimized to establish a convenient and quality method of producing conditional knock-outs in the mouse system.

The Cre-*loxP* system and critical splicing components were key in the functionality of the DECAI intron cassette. The most essential elements of an intron are the 5' splice donor site, branch point and 3' splice acceptor site. The DECAI cassette contains these valuable intronic components and incorporates *loxP* sites in the proper location and orientation to effect gene excision. It is important to note the size of this intron is 201 base pairs prior to the inclusion of DNA sequences for cloning purposes. Compared to other intron methods used, as in the CRISPR-FLIP model which is nearly 2 kb (26), the

DECAI cassette is fairly small in size. These features of the AI made it easy to construct my plasmid and preserve critical components necessary for the function of the intron. Also, I was able to utilize convenient unique restriction sites, Eag1 and Msc1, which allowed for ease of cloning the cassette into my test gene.

The ability of the DECAI intron cassette to be deleted by use of the Cre-lox system makes it a convenient approach. I highlight the importance of the branch point for splicing because without it, splicing may not proceed. Therefore, it is significant that, in this system, the floxed branchpoint is removed to disrupt the splicing process which will facilitate gene inactivation. By testing the functionality of the *loxP* sites within the plasmid by transformation of Cre-expressing bacteria cells into the plasmid, I was able to determine that the AI was capable of Cre-deletion.

Having determined that the *loxP* sites worked led me to conclude it would be a practical system to use for developing a conditional allele. The Neo^R-AI plasmid and Neo^R gene were separately and randomly introduced into mESCs by electroporation, with all intronic components and the Neo^R gene intact. It is important to highlight that after G418 selection for 10 days, the number of colonies carrying the Neo^R-AI plasmid were -4.8-fold fewer than those with of the Neo^R gene. These results suggest that the viability of clones was affected, which suggested gene expression had been reduced in the Neo^R gene due to AI insertion. It is possible that clones that survived G418 selection acquired a higher copy number, and thus more expression, than those that did not survive G418 selection. Assays, such as qPCR could be performed to measure the copy number among clones. Whether this could affect other genes in which the artificial intron is incorporated is unknown. This should be closely assessed when used.

In eukaryotic systems exon-intron splice junctions are important to initiate splicing machinery. At the splice donor site, the 3' exon sequences preferred are C/A, A,G, and 5' intron sequence is G/T. The splice acceptor site, favors A, G at the 3' intron end and G/A at the 5' exon end. If these sequences are not recognized by the spliceosome, splicing typically will not occur. Therefore, the site of insertion of the AI should be properly considered prior to introducing the intron cassette into the target gene.

Although the Neo^R-AI plasmid had not achieved desired clone survival, because these cell colonies did survive G418 selection, it was evident that the Neo^R-AI plasmid DNA had been acquired by the cells. The plasmid DNA was randomly integrated into the mESCs and it is not clear as to where it inserted nor the number of genes integrated with the DNA all of which could have had an effect on my results. In spite of that, once the Neo^R-AI cells were exposed to Cre, the branch point should have been deleted and clone survival should have been reduced as expected.

To assess the efficiency of Cre-mediated branch point removal from this AI construct, a plus/minus screen for Neomycin resistance was implemented on Neo^R-AI Cre^+ cell colonies. My results (7 Neo^S-AI Cre^+ colonies and 41 Neo^R-AI Cre^+ colonies) were unexpected, as I expected to recover only Neo^S-AI Cre deleted colonies. Since all colonies under Neomycin selection came from the same screen, in which they were presumed to have acquired Cre-recombinase through acquisition of the $Hygro^R$-Cre plasmid, I expected that all colonies would be Cre-recombined, resulting in the removal of the branch point. Translational stop codons in the intron should have terminated translation, thus inactivating the gene. Sanger sequencing, however, indicated that the branchpoint had not been deleted in the colonies. The AI remained fully intact, complete with its intronic components and functional Neo^R gene.

It remains unclear as to why Cre-recombinase did not remove the branch point. One possibility is that there may have been a mutation in Cre-recombinase of the Hygro^{R}-Cre plasmid. However, in this case, all cells should have been resistant. A second possibility is that random integration may have led to multiple gene copies in some cells and possibly orienting the loxP sites in various configurations. This could be intriguing, as it could also explain why some cells were Neo^{S} and others Neo^{R}. However, this does not explain why Neo^{S}-AI cell colonies are similar to Neo^{R}-AI cell colonies at the DNA level. Although these two groups responded differently to Neomycin resistance selection, their DNA sequence is the same. A better method to test this system may be to put the Neo^{R}-AI gene as a single copy into a known location. Planning for targeted transgenesis where one copy of the gene is inserted into a known locus such as ROSA26 may minimize insertion of multiple gene copies. The Rosa26 locus has become a favorable site for integration and expression of transgenes and conditional gene expression can be accomplished by using the Cre-loxP system (27).

Another caveat to the system to assess this AI approach chose may simply be due to the Neo^{R} gene itself. The effects of Neomycin B and other aminoglycosides on splicing has been (27). These studies indicated that Neomycin B inhibits splicing indirectly through interference with translation similarly in vivo and in vitro (28). It was also demonstrated that the presence of Neomycin B enhanced cryptic splice sites that were associated with an alternative hairpin structure (28). Cryptic splice sites are present at the mRNA level and once activated, they can interact with the spliceosome and disrupt splicing. Eukaryotes contain increased numbers of cryptic splice sites that more often may not be immediately detectable and are typically suppressed by stronger neighboring splice sites (27-28). There may also be a relationship between the location of cryptic

splice sites and introns; cryptic splice sites in the exons expressed in one species may match the same location of introns in similar genes from another species (29-30). It is possible that where I inserted the DECAI intron cassette, in the presence of Neomycin, may have been more ideal for the activation of cryptic splice sites causing interference with translation or perhaps inhibiting the function of Cre-recombinase.

This project introduced a prokaryotic gene, NeoR, into a eukaryotic system. The NeoR gene is a bacterial gene thereby lacking introns. Manipulation of this gene by directly introducing DNA from a foreign source may have potentially altered cellular performance. Different model systems may vary in requirements to carry out splicing and particular cellular functions. For future testing, it would be helpful to study the best region to place the intron and consider what may compromise the splice site. When contemplating using this AI approach for gene inactivation, one should understand the critical natures for splicing in their system of choice.

At this stage of understanding, I believe the potential for the use of this artificial intron technology to produce conditional alleles in the mouse system is still feasible, however it should be used with caution. A thorough exploration of the gene of interest and developing a test system will offer a more effective platform and approach for its use. The opportune design of the AI supports the fairly straight-forward genetic engineering tool of the Cre-lox system, as well as prospective modifications to the AI to accommodate unique models. Therefore, with additional validation I would expect this approach of generating conditional knock-outs to become translatable across various model systems, leading to further investigations of genetic impacts on biological development and disease.

BIBLIOGRAPHY

1. Hall, B., Limaye, A., & Kulkarni, A. (2009). Overview: generation of gene knockout mice. Current Protocols in Cell Biology, Chapter 19(1), Unit 19.12 19.12.1–19.12.17. https://doi.org/10.1002/0471143030.cb1912s44
2. Iredale, J. (1999). Demystified ... gene knockouts. Molecular Pathology, 52(3), 111–116. https://doi.org/10.1136/mp.52.3.111
3. Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. Genesis (New York, N.Y. : 2000), 26(2), 99–.
4. Kim, H., Kim, M., Im, S., & Fang, S. (2018). Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. Laboratory Animal Research, 34(4), 147–159. https://doi.org/10.5625/lar.2018.34.4.147
5. Sternberg, N., & Hamilton, D. (1981). Bacteriophage P1 site-specific recombination: I. Recombination between loxP sites. Journal of Molecular Biology, 150(4), 467–486. https://doi.org/10.1016/0022-2836(81)90375-2
6. Hamilton, D., & Abremski, K. (1984). Site-specific recombination by the bacteriophage P1 lox-Cre system : Cre-mediated synapsis of two lox sites. Journal of Molecular Biology, 178(2), 481–486. https://doi.org/10.1016/0022-2836(84)90154-2
7. Branda, C., & Dymecki, S. (2004). Talking about a Revolution: The Impact of Site-Specific Recombinases on Genetic Analyses in Mice [Review of Talking about a Revolution: The Impact of Site-Specific Recombinases on Genetic Analyses in Mice]. Developmental Cell, 6(1), 7–28. Elsevier Inc. https://doi.org/10.1016/S1534-5807(03)00399-X

8. Renninger, Sabine & Schonthaler, Helia & Neuhauss, Stephan & Dahm, Ralf. (2011). Investigating the genetics of visual processing, function and behaviour in zebrafish. Neurogenetics. 12. 97-116. 10.1007/s10048-011-0273-x.
9. Alberts B, Johnson A, Lewis J, K. Roberts and P. Walter (eds). Molecular Biology of the Cell. 4th edition. New York: Garland Science; 2002. From DNA to RNA. (Also used for FIG 2 and 4).
10. Citation: Clancy, S. (2008) RNA splicing: introns, exons and spliceosome. *Nature Education* 1(1):31
11. Wachtel, C., & Manley, J. (2009). Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. Molecular bioSystems, 5(4), 311–. https://doi.org/10.1039/b820828j
12. Black, D. (2003). MECHANISMS OF ALTERNATIVE PRE-MESSENGER RNA SPLICING. Annual Review of Biochemistry, 72(1), 291–336. https://doi.org/10.1146/annurev.biochem.72.121801.161720 L
13. Sheth, N., Roca, X., Hastings, M., Roeder, T., Krainer, A., & Sachidanandam, R. (2006). Comprehensive splice-site analysis using comparative genomics. Nucleic Acids Research, 34(14), 3955–3967. https://doi.org/10.1093/nar/gkl556
14. Russell, P. (2013). iGenetics: A Molecular Approach. In iGenetics. Pearson Education UK.
15. Simpson, C., Clark, G., Davidson, D., Smith, P., & Brown, J. (1996). Mutation of putative branchpoint consensus sequences in plant introns reduces splicing efficiency. *The Plant Journal*, *9*(3), 369–380. https://doi.org/10.1046/j.1365-313X.1996.09030369.x

16. Taggart, A., Lin, C., Shrestha, B., Heintzelman, C., Kim, S., & Fairbrother, W. (2017). Large-scale analysis of branchpoint usage across species and cell lines. Genome Research, 27(4), 639–649. https://doi.org/10.1101/gr.202820.11.
17. Woll M.G., Naryshkin N.A., Karp G.M. (2017) Drugging Pre-mRNA Splicing. In: Garner A. (eds) RNA Therapeutics. Topics in Medicinal Chemistry, vol 27. Springer, Cham. https://doi.org/10.1007/7355_2017_12.
18. Brinster, R., Allen, J., Behringer, R., Gelinas, R., & Palmiter, R. (1988). Introns increase transcriptional efficiency in transgenic mice. Proceedings of the National Academy of Sciences of the United States of America, 85(3), 836–840. https://doi.org/10.1073/pnas.85.3.836
19. Choi, T., Huang, M., Gorman, C., & Jaenisch, R. (1991). A generic intron increases gene expression in transgenic mice. Molecular and Cellular Biology, 11(6), 3070–3074. https://doi.org/10.1128/mcb.11.6.3070
20. Economides, A., Frendewey, D., Yang, P., Dominguez, M., Dore, A., Lobov, I., Persaud, T., Rojas, J., McClain, J., Lengyel, P., Droguett, G., Chernomorsky, R., Stevens, S., Auerbach, W., DeChiara, T., Pouyemirou, W., Cruz, J., Feeley, K., Mellis, I., Yancopoulos, G. (2013). Conditionals by inversion provide a universal method for the generation of conditional alleles. Proceedings of the National Academy of Sciences - PNAS, 110(34), E3179–E3188. https://doi.org/10.1073/pnas.1217812110
21. Guzzardo, P., Rashkova, C., Dos Santos, R., Tehrani, R., Collin, P., & Bürckstümmer, T. (2017). A small cassette enables conditional gene inactivation by CRISPR/Cas9. Scientific Reports, 7(1), 16770–11. https://doi.org/10.1038/s41598-017-16931-z

22. Labun, K., Montague, T. G., Krause, M., Torres Cleuren, Y. N., Tjeldnes, H., & Valen, E. CHOPCHOP v3: expanding the CRISPR web toolbox beyond genome editing. Nucleic Acids Research (2019).

23. Valera, A., Perales, J., Hatzoglou, M., & Bosch, F. (1994). Expression of the neomycin-resistance (neo) gene induces alterations in gene expression and metabolism. Human Gene Therapy, 5(4), 449–456. https://doi.org/10.1089/hum.1994.5.4-449.

24. Lewandoski, M. (2001). Conditional control of gene expression in the mouse. Nature Reviews. Genetics, 2(10), 743–755. https://doi.org/10.1038/35093537

25. Andersson-Rolf, A., Mustata, R., Merenda, A., Kim, J., Perera, S., Grego, T., Andrews, K., Tremble, K., Silva, J., Fink, J., Skarnes, W., & Koo, B. (2017). One-step generation of conditional and reversible gene knockouts. Nature Methods, 14(3), 287–289. https://doi.org/10.1038/nmeth.4156

26. Bouabe, H., & Okkenhaug, K. (2013). Gene Targeting in Mice: A Review. In Methods in molecular biology (Clifton, N.J.) (Vol. 1064, pp. 315–336). Humana Press. https://doi.org/10.1007/978-1-62703-601-6_23

27. Waldsich, C., Semrad, K., & Schroeder, R. (1998). Neomycin B inhibits splicing of the td intron indirectly by interfering with translation and enhances missplicing in vivo. *RNA (New York, N.Y.)*, *4*(12), 1653–1663. https://doi.org/10.1017/s1355838298981444

28. Kapustin, Y., Chan, E., Sarkar, R., Wong, F., Vorechovsky, I., Winston, R. M., Tatusova, T., & Dibb, N. J. (2011). Cryptic splice sites and split genes. *Nucleic acids research*, *39*(14), 5837–5844. https://doi.org/10.1093/nar/gkr203

29. Padgett, R. A., Grabowski, P. J., Konarska, M. M., Seiler, S., & Sharp, P. A. (1986). Splicing of messenger RNA precursors. *Annual review of biochemistry*, *55*, 1119–1150. https://doi.org/10.1146/annurev.bi.55.070186.005351

30. Green M. R. (1986). Pre-mRNA splicing. *Annual review of genetics*, *20*, 671–708. https://doi.org/10.1146/annurev.ge.20.120186.003323

VITA

Amber Thomas-Gordon was born in Port Arthur, Texas, the daughter of Julie A. Moore and Burnette D. Pace. After completing her work at Ross Sterling High School, Baytown, Texas in 2002, she entered Prairie View A&M University in Prairie View, Texas. She received the degree of Bachelor of Science with a major in Biology in May 2007. In August of 2017, she entered The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences.

Permanent Address:

1800 Barker Cypress Road, 406

Houston, TX. 77084