Video Article

In vivo Application of the REMOTE-control System for the Manipulation of Endogenous Gene Expression

Nicole A. Vander Schaaf[1], Shirley Oghamian[2], Jin-A Park[1], Liang Kang[1], Peter W. Laird[1], Kwang-Ho Lee[1]

[1]Center for Epigenetics, Van Andel Research Institute

[2]Amgen

Correspondence to: Kwang-Ho Lee at KwangHo.Lee@vai.org

Abstract

Here we describe a protocol for implementing the REMOTE-control system (*Re*versible *M*anipulation *o*f *T*ranscription at *E*ndogenous loci), which allows for reversible and tunable expression control of an endogenous gene of interest in living model systems. The REMOTE-control system employs enhanced *lac* repression and *tet* activation systems to achieve down- or upregulation of a target gene within a single biological system. Tight repression can be achieved from repressor binding sites flexibly located far downstream of a transcription start site by inhibiting transcription elongation. Robust upregulation can be attained by enhancing the transcription of an endogenous gene by targeting tet transcriptional activators to the cognate promoter. This reversible and tunable expression control can be applied and withdrawn repeatedly in organisms. The potency and versatility of the system, as demonstrated for endogenous *Dnmt1* here, will allow more precise in vivo functional analyses by enabling investigation of gene function at various expression levels and by testing the reversibility of a phenotype.

Video Link

The video component of this article can be found at https://www.jove.com/video/59235/

Introduction

Genetic knockout or transgenic approaches have been effective means to study gene function in animal models. However, expression regulation by these approaches is dichotomous (on/off), non-temporal, and thus is not capable of revealing the complete functional spectrum of a gene. Conditional Cre/L*oxP* technologies have allowed spatio-temporal inactivation or activation of gene function, but their dichotomous nature continues to pose limitations, such as cell lethality and irreversibility[1,2,3]. In order to fill this void, conditional knockdown approaches have been developed using *tet*-regulated shRNA or miRNA[4]. However, off-target effects remain a concern for RNAi[5] and have been challenging to control in vivo. More recently, CRISPR/Cas-mediated transcriptional-control technologies have introduced a more versatile approach to achieving both up- and downregulation of endogenous gene expression and demonstrated their utilities[6,7]. However, the effectiveness of CRISPR/Cas-mediated transcriptional control is as yet unclear in vivo, and the reversibility of KRAB-based repression remains to be seen, as strong repression by KRAB and its interacting protein KAP1 has been shown to induce permanent gene silencing[8,9].

In order to address these limitations, we have developed a novel transcriptional regulatory system capable of conditionally controlling endogenous gene expression in a reversible and tunable manner in mice using engineered prokaryotic binary transcriptional regulatory systems[10]. Prokaryotic binary transcriptional regulatory systems with regulatory ligands, such as *lac* and *tet*, have enabled such reversible and tunable expression control[11,12,13,14]. However, the inadequate repression potency of the current binary systems has impeded their broad adoption for controlling endogenous gene expression in mammals. We developed an enhanced lac repression system sufficiently potent for the repression of endogenous genes and employed a novel strategy of targeting *tet* transcriptional activators directly to the cognate promoter of an endogenous gene to achieve robust upregulation (**Figure 1**)[10]. With this technology, we have achieved nearly two orders of magnitude expression control of the endogenous *Dnmt1* gene in a tunable, inducible, and reversible manner[10]. Here we provide step-by-step instructions for its in vivo application to other genes and organisms using mice as a model species.
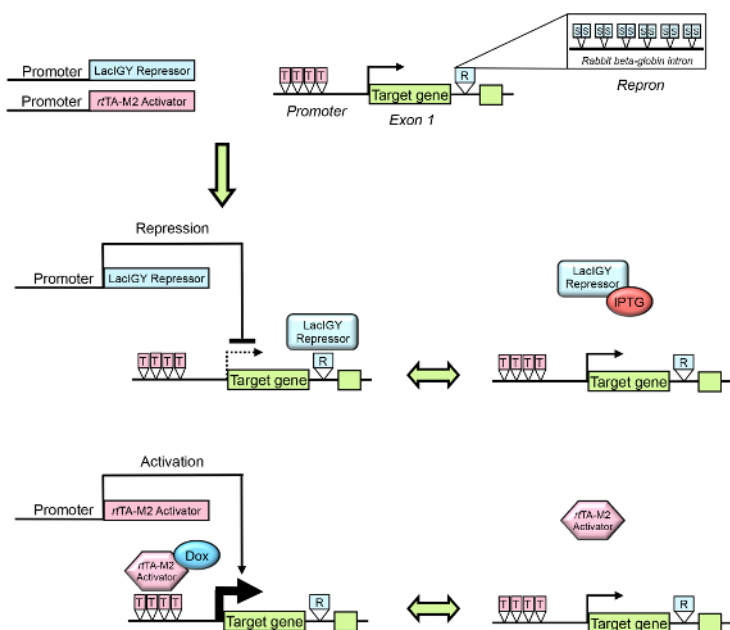
Figure 1: Overview of the REMOTE-control system. The transcription of an endogenous target gene can be regulated using engineered *lac* repressor and *tet* activator systems. The target gene promoter or intron is engineered to contain operators for the tight-binding LacIGY repressor and/or the *rt*TA-M2 activator. R indicates Repron (*Rep*ression int*ron*), which contains 12 symmetric lac operators (S) plus a partial rabbit beta-globin intron. T indicates *tet* operator. The repressor and/or activator is/are expressed from a tissue-specific promoter. The expression of the target gene can then be reversibly tuned to the desired expression level by administration of IPTG (isopropyl β-D-1-thiogalactopyranoside, an antagonist of the LacIGY repressor) or Doxycycline (Dox). This figure has been modified from Lee et al.[10] Please click here to view a larger version of this figure.

Before beginning this protocol, review **Table 1** to identify the relevant steps for the desired control of gene expression. For example, to engineer a mouse that enables reversible downregulation of "Gene X", complete sections 1, 3, and 4 of the below protocol. **Table 1** also summarizes the needed components of the REMOTE-control system.

Desired Expression Change	Repression only	Activation only	Both Repression and Activation
Relevant Sections of Protocol	1, 3-4	2-4	1-4
REMOTE-control Sequence Needed in Target Gene	Repron ("Repression intron"; 12 symmetric lac operators plus a partial rabbit beta-globin intron)	Tet operator(s)	Repron and Tet operator(s)
REMOTE-control Sequence Location	Intron	Promoter	Intron & Promoter
Activator/Repressor Needed for Desired Control	LacIGY Repressor	rtTA-M2 Activator	LacIGY Repressor and rtTA-M2 Activator
Regulatory Ligands	IPTG	Doxycycline	IPTG and/or Doxycycline

Table 1: Overview of REMOTE-control components.

Protocol

All animal procedures were conducted with the approval of the Institutional Animal Care and Use Committee (IACUC) of the University of Southern California and the Van Andel Research Institute and in compliance with the Guide for the Care and Use of Laboratory Animals from the National Institutes of Health[15].

1. Modify the gene of interest for repression by REMOTE-control

1. Using the guidelines below, identify a transcriptionally inert intron toward the 5' end of the gene of interest for insertion of the Repron sequence ("*Rep*ression int*ron*"; 12 symmetric lac operators plus a partial rabbit beta-globin intron). Be aware of alternative promoters for the gene of interest, and choose an intron according to the transcript(s) to be controlled (i.e. an intron shared by all transcripts or one specific to a desired transcript).

NOTE: For genes without an intron or with ones toward the 3' end, insert the Repron sequence into the promoter (see Lee, et al.[10] for a detailed procedure).

1. Obtain the genomic sequence for the gene of interest.
   1. Navigate to the UCSC Genome Browser[16,17], and select the latest draft of the mouse genome (Mouse GRCm38/mm10 at time of publication), which is currently under the **Genomes** tab.
   2. Enter the name or symbol of the gene of interest into the search bar to view the transcripts for the gene. Click **go**.
   3. Select the desired transcript variant for the gene of interest.
   4. Click on the gene symbol next to the transcript variant of interest (the symbol of the previously selected transcript will be in a dark box).
   5. Under the **Sequence and Links to Tools and Databases** banner, click the **Genomic Sequence** link.
   6. For **Sequence Retrieval Region Options**, select only Exons (5' UTR, CDS, and 3' UTR), Introns, and the default **One FASTA record per gene**. For **Sequence Formatting Options**, select **Exons in upper case, everything else in lower case** and **Mask Repeats to N** (to conceal repetitive sequences). Then click **submit**.
   7. Save this sequence, preserving the upper- and lowercase formatting, in a document or program that can be annotated.

2. Avoid interruption of CpG islands, which may indicate regions with gene regulatory function.
   NOTE: Though a 5' intron is preferable for insertion of the Repron sequence, the first intron may contain transcriptional regulatory elements such as CpG islands (examined in this step) or enhancer elements (examined in step 1.1.4), which are to be avoided for Repron insertion.
   1. Under the **Expression and Regulation** banner of the UCSC Genome Browser, select **show** for the **CpG Islands** track, and click **refresh**.
   2. Zooming in on the 5' introns, click on each CpG island (shown in green at the time of publication if present), and select **View DNA for this feature**.
   3. After selecting **Mask repeats to N**, select **get DNA** to obtain the CpG island sequence.
   4. Overlay these sequences with the original sequence file, and annotate these as intronic regions to avoid due to possible gene regulatory functions.

3. Avoid interruption of non-coding RNAs, in case they have a regulatory function.
   1. Under the **Genes and Gene Predictions** banner of the UCSC Genome Browser, select **show** for the **GENCODE (Ensembl)** track, and click **refresh**.
   2. Zooming in on the 5' introns, click on each non-coding RNA (shown in green at the time of publication if present), and obtain the DNA sequence for each by clicking on its chromosomal coordinates.
   3. In the **View** dropdown menu, select **DNA**, and click **Mask repeats to N**. Then click **get DNA**.
   4. Overlay these sequences with the original sequence file, and annotate these as intronic regions to avoid due to possible gene regulatory functions.

4. Avoid intronic regions with enhancer signatures in the tissue(s) of interest, such as H3K4me1, H3K27ac, and DNase I hypersensitivity[18,19,20,21], as well as CTCF binding sites, which regulate enhancer looping[22,23].
   1. Navigate to the ENCODE database[24,25], and select the **Experiments** icon.
   2. For the assay type, select **ChIP-seq** or **DNase-seq**, and populate the other categories (**Organism**, **Biosample type**, etc.) according to the cells to be engineered.
   3. After feature selection, hover over the pictograms in blue, and select the one for which **View Results as List** appears (left-most pictogram at time of publication).
   4. Select the datasets for the targets of H3K4me1, H3K27ac, DNase I, and CTCF that most closely match the cells to be engineered.
   5. Within each relevant dataset, scroll to the **Files** section, verify that mm10 (or the latest genome version) and UCSC are selected, and click the **Visualize** button.
   6. Now in the UCSC Genome Browser and zooming in on the 5' introns for the gene of interest, click on each H3K4me1, H3K27ac, DNase I, and CTCF peak in their respective annotated peak tracks.
   7. Obtain the DNA sequence for each peak region by clicking on its chromosomal coordinates.
   8. In the **View** dropdown menu, select **DNA**, and click **Mask repeats to N**. Then click **get DNA**.
   9. Overlay these sequences with the original sequence file, and annotate these as intronic regions to avoid due to possible gene regulatory functions.

5. Avoid disruption of consensus splicing sequences, so as not to interfere with appropriate splicing of the exons.
   NOTE: Although sequences required for splicing are largely confined to about six bases at the 5' end of an intron[26] and about 60 bases at the 3' end[27], it is advisable to select a large intron, allowing for considerably wider margins from these consensus regions to minimize the likelihood of impacting splicing. Use of intron analysis tools, such as SVM-BPfinder[28], is recommended for a more thorough analysis of potential splice sequences.

2. After identifying an intronic region that meets the above criteria (as exemplified for *Dnmt1* in the supplementary data), screen the sequence using an online sgRNA design tool to identify an sgRNA in the region with high specificity and predicted efficiency scores.
   1. Navigate to an online sgRNA design tool of choice, such as CRISPOR[29].
   2. Enter the sequence of the intronic region of interest, specify the relevant reference genome, and select the desired Protospacer Adjacent Motif (PAM). Click **Submit**.
   3. Sort the predicted sgRNAs by specificity score, and select one or more sgRNAs that also have a high predicted efficiency score[29].
      1. Optional: To maximize the likelihood of using an effective sgRNA, first test the cleavage efficiency of several top-scoring sgRNAs in an in vitro assay[30], and proceed with the most efficient sgRNA in vivo.

3. Design a DNA template containing a PITT (Pronuclear Injection-based Targeted Transgenesis) landing pad sequence, as exemplified in **Supplementary Figure 1A,B**, flanked on both sides by 60-base homology arms that correspond to the sgRNA cut site[31].
   NOTE: The landing pad contains two heterotypic *loxP* sites (JT15 and Lox2272) and enables targeted insertion of large sequences through a two-step approach; be sure that the junctions of this insertion do not create cryptic splice sites. Alternatively, an embryonic stem cell (ESC)-based knock-in strategy can be used to insert the Repron sequence directly.
4. Prepare the sgRNA, Cas9 protein, and the single-stranded DNA (ssDNA) template for the landing pad, and microinject into fertilized eggs (from B6C3F1/J mice or other desired strain) according to established protocols[32,33].
5. Screen for mice with the landing pad knock-in.
   1. Design PCR primers complementary to the genomic locus but outside of the regions targeted by the homology arms, as demonstrated for *Dnmt1* in **Supplementary Figure 1C**. Avoid repetitive genomic sequences when designing the primers.
   2. Extract DNA from tail clips of the mice according to established protocols[34].
   3. Use PCR and gel electrophoresis to identify mice with a landing pad insertion.
   4. Confirm that the knock-in was successful by sequencing the PCR products.
      NOTE: Unwanted large deletions and rearrangements can be introduced by CRISPR/Cas9[35], so careful screening for off-target editing is advised before proceeding[35,36,37].
6. Using fertilized eggs from the landing pad mice, microinject iCre mRNA[38] and a transgenic plasmid containing the Repron sequence flanked by JTZ17 and Lox2272 recombination sites[31,39,40,41] according to established methods[38,42,43].
7. Screen for mice with the Repron insertion.
   1. Design PCR primers complementary to the genomic locus, outside of the Repron insert. Avoid repetitive genomic sequences when designing the primers.
   2. Extract DNA from tail clips of the mice according to established protocols[34].
   3. Use PCR and gel electrophoresis to identify mice with a Repron insertion.
   4. Confirm the knock-in was successful by sequencing the PCR products.

2. Modify the gene of interest for upregulation by REMOTE-control

1. Using the guidelines below, identify a region in the promoter of the gene of interest that is unlikely to perturb promoter function upon insertion of *tet* operator sequences. Be aware of alternative promoters for the gene of interest, and choose a promoter according to the transcript(s) to be controlled (i.e. a promoter shared by all transcripts or one specific to a desired transcript).
   1. Obtain the genomic sequence for the promoter of interest.
      1. Navigate to the UCSC Genome Browser[16,17], and select the latest draft of the mouse genome (Mouse GRCm38/mm10 at time of publication), which is currently under the **Genomes** tab.
      2. Enter the name or symbol of the gene of interest into the search bar to view the transcripts for the gene. Click **go**.
      3. Select the desired transcript variant for the gene of interest.
      4. Click on the gene symbol next to the transcript variant of interest (the gene symbol of the previously selected transcript will be in a dark box).
      5. Under the **Sequence and Links to Tools and Databases** banner, click the **Genomic Sequence** link.
      6. For **Sequence Retrieval Region Options**, select only **Promoter/Upstream by 1000 bases**. For **Sequence Formatting Options**, select **Mask Repeats to N** (to conceal repetitive sequences). Then click **submit**.
      7. Save this promoter sequence in a document or program that can be annotated.
   2. Select regions that are free of putative transcription factor binding sites, as interrupting these sequences may alter endogenous gene expression.
      1. Navigate to the UCSC Genome Browser, and open the latest version of the mouse genome.
      2. Above the **Mapping and Sequencing** banner, select **track hubs**.
      3. Click **mm10** next to the hub name of **JASPAR 2018 TFBS** (or the latest JASPAR version).
      4. Enter the name or symbol of the gene of interest into the search bar to view the transcripts for the gene. Click **go**.
      5. Select the desired transcript variant for the gene of interest.
      6. Scroll down to the **JASPAR 2018 TFBS** banner, and click the dropdown arrow to select the desired viewing option for the track, such as **squish**. Click **refresh**.
      7. Zoom into the promoter region of the gene of interest, and identify regions that are free (or relatively free) of transcription factor binding sites according to the JASPAR track. Record the chromosomal coordinates of these regions.
      8. Obtain the genomic sequence of these chromosomal coordinates by typing them into the search bar and clicking **go**. In the **View** dropdown menu, select **DNA**, and click **Mask repeats to N**. Then click **get DNA**.
      9. Overlay these sequences with the original sequence file, and annotate these as ideal promoter regions to target due to lack of transcription factor binding.
   3. Within these sequences, select a perturbable promoter region that is upstream but near the transcription start site of the gene of interest.
      NOTE: An insertion that is too close to the transcription start site may increase the chance of impairing promoter activity, but an insertion that is too far may decrease the level of upregulation. Inserting *tet* operator sequences about 200 basepairs upstream of the transcription start site has resulted in robust upregulation of the two promoters tested (see discussion)[10].
2. Screen the selected promoter region using an online sgRNA design tool to identify an sgRNA in the region with high specificity and predicted efficiency scores.
   1. Navigate to an online sgRNA design tool of choice, such as CRISPOR[29].

2. Enter the sequence of the region of interest, specify the relevant reference genome, and select the desired Protospacer Adjacent Motif (PAM). Click **Submit**.
3. Sort the predicted sgRNAs by specificity score, and select one or more sgRNAs that also have a high predicted efficiency score[29].
    1. Optional: To maximize the likelihood of using an effective sgRNA, first test the cleavage efficiency of several sgRNAs in an in vitro assay[30], and proceed with the most efficient sgRNA in vivo.

3. Design a DNA template containing *tet* operators flanked on both sides by 60-base homology arms that correspond to the sgRNA cut site[31,33].
    NOTE: The number of *tet* operators is customizable; insertion of two to four *tet* operator sequences in tandem has previously been shown to be effective, but in principle more operators are desirable for driving higher expression. Alternatively, an ESC-based knock-in strategy can be used to insert the *tet* operators.
    1. Optional: For experimental evidence that the proposed modifications will likely not disrupt the endogenous transcriptional activity of the gene of interest, clone the engineered promoter sequence into a firefly luciferase vector (such as pGL3-Basic), and compare its efficacy to the original promoter using a luciferase assay.

4. Prepare the sgRNA, Cas9 protein, and the ssDNA template containing the *tet* operator sequences, and microinject into fertilized eggs (from B6C3F1/J mice or other desired strain) according to established protocols[32,33].
    NOTE: Due to the complexity of synthesizing a repetitive sequence, an in vitro transcription/reverse transcription-based approach is recommended to synthesize a ssDNA template from a double-stranded DNA (dsDNA) plasmid[44]. Alternatively, a dsDNA template may be used for microinjection, but the knock-in efficiency may be reduced[45].
5. Screen for mice with knock-in of the *tet* operators.
    1. Design PCR primers complementary to the genomic locus, outside of the regions targeted by the homology arms. Avoid repetitive genomic sequences when designing the primers.
    2. Extract DNA from tail clips of the mice according to established protocols[34].
    3. Use PCR and gel electrophoresis to identify mice with the insertion of the *tet* operators.
    4. Confirm that the knock-in was successful by sequencing the PCR products.
        NOTE: Unwanted large deletions and rearrangements can be introduced by CRISPR/Cas9[35], so careful screening for off-target editing is advised before proceeding[35,36,37].

3. Develop activator- and/or repressor-expressing mice

1. Identify a robustly expressing promoter for the tissue(s) or cell type(s) of interest.
    NOTE: A literature search and the Tissue-Specific Promoter Database[46] may be useful for identifying such a promoter.
2. Place the provided enhanced *lac* repressor or *tet* activator sequence downstream of the desired promoter to generate a transgenic construct.
3. Produce the transgenic mouse line(s) using standard transgenic procedures[42,43,47]. Alternatively, a site-specific transgenesis approach can be used to avoid position effects and to allow single-copy transgene insertion[31,48].
4. Propagate the founders, and determine the expression pattern and level of the transgene in the offspring of each founder. Select lines with robust expression in the intended tissue type(s) for further breeding.
    1. Breed the founders to wildtype mice.
    2. Design PCR primers that will detect the insertion of the activator and/or repressor.
    3. Extract DNA from tail clips of the offspring according to established protocols[34].
    4. Use PCR and gel electrophoresis to identify mice with the insertion.
    5. Confirm the transgene insertion by sequencing of the PCR products.
    6. Propogate the transgenic lines.
    7. Sacrifice a few pups from each line, and collect the tissues of interest for qRT-PCR, immunohistochemistry, and/or western blotting to analyze the expression of the gene/protein of interest in the target tissue type(s).
    8. Select the lines with the strongest expression in the tissues of interest for use in section 4.

4. Manipulate gene expression in vivo

1. Breed the mice with the modified gene of interest (section 1 and/or 2) with the transgenic mice from section 3 according to established breeding practices[49]. For maximal expression control, breed the mice to homozygosity for the modified allele.
2. For reversion or adjustment of repression of the target gene, administer IPTG in the drinking water.
    1. To experimentally determine the dose of IPTG to use, treat mice of the appropriate genotype and controls with one of a range of doses (recommended starting range: 0 – 400 mM IPTG) for at least one week[50,51]. Include at least three mice per treatment group, and select the age and sex of mice that are most relevant to the planned future experiments. NOTE: Breeding pairs of mice can be treated with IPTG water to provide developmental exposure of IPTG to the offspring if desired[13], or mice can begin treatment any time after birth.
        1. Dissolve the desired mass of IPTG in sterile distilled water on the day of administration, and stir with a stir bar for 5 minutes or until fully dissolved.
        2. Wrap the bottle with foil, and administer the IPTG water in a light-protected bottle. Replace twice a week. Provide the same source of water to mice receiving 0 mM IPTG.
        3. After at least one week, sacrifice the mice, and analyze the expression of the gene of interest in the target tissue(s) using qRT-PCR, immunohistochemistry, and/or western blotting.

    4. Identify the dose that restores the expression of the gene of interest to that of wildtype controls, and use this dose to achieve normal expression of the gene in future experiments.

3. For induction of gene upregulation, administer Doxycycline (Dox) in the diet.
   1. To experimentally determine the concentration of Dox to administer, treat mice of the appropriate genotype and controls with one of a range of Dox concentrations (recommended starting range: 0 – 5000 mg/kg Doxycycline Hyclate) for at least one week[52,53], including at least three mice per treatment group. Purchase Dox-containing mouse food from a commercial vendor.
   NOTE: Breeding pairs of mice can be treated with Dox food to provide developmental exposure of Dox to the offspring if desired[54,55], or mice can begin treatment any time after birth.
   2. Replace the diet once per week. Provide the same base diet to mice receiving Dox-free food.
   3. After at least one week, sacrifice the mice, and analyze the expression of the gene of interest in the target tissue(s) using qRT-PCR, immunohistochemistry, and/or western blotting.
   4. Identify the dose that elevates the expression of the gene of interest to the desired level, and use this dose to achieve overexpression of the gene in future experiments.

Representative Results

The repression capability of the REMOTE-control system has been demonstrated in two different approaches thus far. In the first approach, *lac* repressor binding sites were inserted at the endogenous promoter of the *Dnmt1* gene. In the second approach, which is recommended by this protocol, the repressor binding sites were inserted into a downstream intron to avoid the potential risk of affecting promoter function by the insertion and thereby to simplify application of the REMOTE-control system. Both approaches resulted in successful repression (**Figure 2A,B** and **Figure 3A-C**)[10]. *Dnmt1* expression was repressed to 15% of the unregulated levels using the promoter-based approach (**Figure 2A**). This tight repression was reversed in a dose-dependent manner by treating mice with varying amounts of IPTG (**Figure 2A**). The observed *Dnmt1* repression was validated at the protein level by immunostaining (**Figure 2B**). We did not observe any noticeable difference in *Dnmt1* expression between *Dnmt1*$^{+/+}$ and *Dnmt1*$^{LO/LO}$ mice, confirming that our *lac* operator insertion had not disrupted normal promoter function[10]. The intron-based approach achieved more than 90% repression from operators located several kilobases downstream of the transcription start site by attenuating transcription elongation (**Figure 3A, B**)[10]. This intron-based approach was further validated on seven additional robust promoters (**Figure 3C**). Invariably tight repression was achieved from all of the promoters tested. No correlation between the residual expression levels and the strengths of the promoters was observed, suggesting that the repression capacity of our intron-based repression system exceeds the transcriptional potency of all of the robust promoters we tested (**Figure 3C**).

The in vivo upregulation capability of the REMOTE-control system was also demonstrated on the *Dnmt1* gene. We introduced two copies of the *tet* operator into the *Dnmt1* promoter, together with *lac* operator sequences, to allow for either upregulation or downregulation depending on which effector protein is present. Robust upregulation and downregulation of *Dnmt1* expression, close to two orders of magnitude (10% to 650%), were achieved in ESCs containing the modified endogenous *Dnmt1* allele (*Dnmt1LGT*) (**Figure 4A**)[10]. Both regulations were fully reversible and inducible by IPTG and Dox treatments, respectively (**Figure 4A**). We next introduced the *Dnmt1LGT* modification into the mouse germline to test the in vivo upregulation capability of the REMOTE-control system. Strong upregulation of *Dnmt1* was observed from the liver, spleen, and kidney, whereas no detectable upregulation in the heart was observed (**Figure 4B**)[7]. The cell cycle-dependent expression pattern of *Dnmt1* and the scarcity of proliferative cells in the heart may underlie this observation[10,56]. It remains to be seen whether this limitation can be overcome by increasing the expression level of the activator or the number of its binding sites.
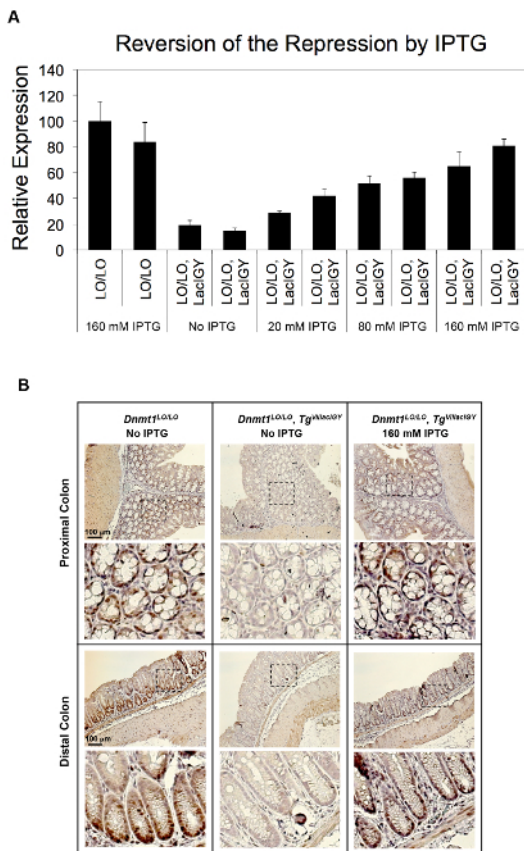
Figure 2: In vivo repression of *Dnmt1* by the LacIGY repressor. (**A**) Mice with *lac* operators (LO) inserted into the *Dnmt1* promoter, with or without expression of *LacIGY*, were treated with various doses of IPTG. qRT-PCR analysis of *Dnmt1* expression shows the dose-dependent reversal of *Dnmt1* repression in vivo by IPTG treatment. Each bar represents data from a different mouse. Data represent mean ± SEM (*n* = 3). (**B**) Immunostaining of Dnmt1 protein in colonic crypts of mice provided drinking water with or without 160 mM IPTG for 3 weeks. This figure has been modified from Lee et al.[10] Please click here to view a larger version of this figure.
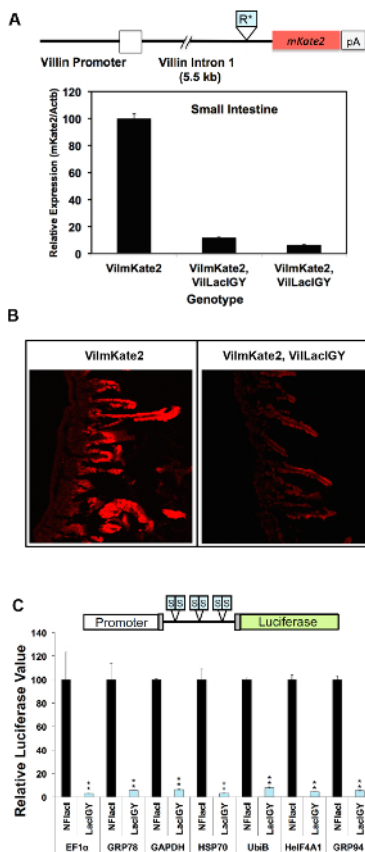
Figure 3: In vivo and in vitro repression of various promoters by the REMOTE-control system. (**A**) An early version of the Repron sequence (R*) was inserted into an intron downstream of the *Villin* promoter in a *Villin-mKate2* transgenic mouse (*VilmKate2*). qRT-PCR analysis of *mKate2* expression in the small intestine of mice with or without the *LacIGY* repressor is shown. Each bar represents data from a different mouse. (**B**) Confocal mKate2 images of the small intestine with and without *LacIGY* expression. (**C**) Six symmetric lac operators (S) were inserted between various promoters and a luciferase reporter. Reporters (50 ng/well in 96-well plate) and repressor plasmids were transiently introduced into NIH/3T3 cells in a 1:1 molar ratio. Luciferase values were assessed 24 h after transfection. These in vitro data represent the percent of luciferase expression in *LacIGY*-expressing cells relative to those expressing non-functional *LacI* (*NFlacI*). *T-tests* were used to determine statistical significance. Data represent mean ± SEM (*n* = 3). *$P \le 0.05$, **$P \le 0.01$. This figure has been modified from Lee et al.[10]
Please click here to view a larger version of this figure.
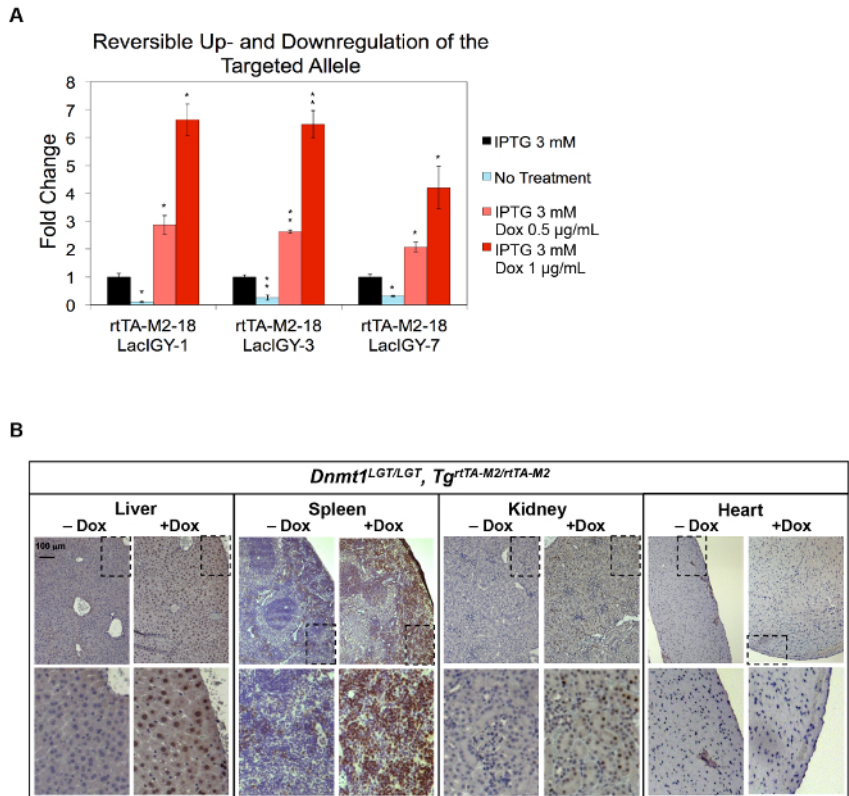
Figure 4: Down- and/or upregulation of *Dnmt1* expression in vitro and in vivo. (A) The complete REMOTE-control system was engineered in cultured ESCs by gene targeting and electroporation approaches. Maximal repression of *Dnmt1* expression was achieved with no treatment while maximal activation was achieved by both IPTG and Dox treatment. Data represent mean ± SEM (*n* = 3). *$P \leq 0.05$, **$P \leq 0.01$ (Welch's *t*-tests). **(B)** In vivo activation of *Dnmt1* by the REMOTE-control system, as demonstrated by immunostaining of Dnmt1 protein in various tissues from REMOTE-control mice. The *LGT* allele represents promoter modification of *Dnmt1* to contain *lac* operator and *tet* activator binding sites. Mice were treated with a normal or Dox-containing diet (5000 mg/kg Doxycycline Hyclate) for one month. This figure has been modified from Lee et al.[10] Please click here to view a larger version of this figure.
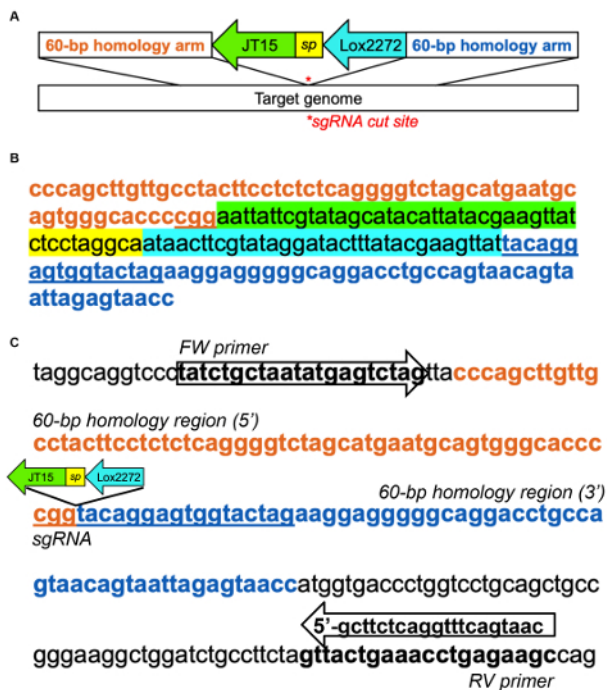
Supplementary Figure 1: Example of landing pad insertion into murine *Dnmt1* intron 1. (**A**) Schematic of DNA template for landing pad insertion, adapted from Quadros et al. (2015)[31]. Heterotypic *loxP* sites, JT15 and Lox2272, are separated by a short spacer sequence (*sp*) and flanked on each side by 60-bp of DNA that is homologous to the target genomic region. (**B**) Sample DNA template for landing pad insertion into the *Dnmt1* intron using the following sgRNA: CTAGTACCACTCCTGTACCG (which targets the reverse strand). The selected intronic region was bioinformatically informed by step 1.1, and the sgRNA was identified using CRISPOR[29]. (**C**) Example of PCR primer design for assessing insertion of the landing pad. PCR primers were designed outside of the homology arms of the template to confirm integration into endogenous *Dnmt1*. The wildtype PCR amplicon is 213 bp; upon insertion, it becomes 291 bp. Please click here to view a larger version of this figure.

Discussion

A critical step and potential limitation of the REMOTE-control system is the challenge associated with the insertion of the repressor and/or activator binding sites without affecting target gene expression. Our original repression approach, as applied to the *Dnmt1* gene, involved insertion of *lac* repressor binding sites within transcriptionally critical regions of a promoter. In order to reduce the risk of affecting promoter function and thus to improve the general applicability of the REMOTE-control system, we developed an intron-based repression approach. The potency of our enhanced *lac* system allowed us to tightly repress the transcription of all the strong promoters we tested at operators located hundreds to several kilobases downstream of the transcription start sites (**Figure 3A–C**)[10]. Importantly, the levels of repression were independent of the transcriptional strengths of the promoters (**Figure 3A–C**)[10]. This suggests that the repression capacity of our intron-based repression system exceeds the transcriptional strength of the tested promoters. In this intron-based approach, it is likely that the repression is mediated through physical interference between two components, the transcription elongation machinery and the lac repressors[57]. This simple repression mechanism and the demonstrated robustness of the intron-based method may render this approach generally applicable to different genes, tissues, and organisms.

The upregulation by the REMOTE-control system requires the transactivator binding sequences to be in proximity to the target gene promoter, which entails a risk of affecting promoter function. However, we found that the position of binding sequences can be outside of the transcriptionally critical region. Both *Dnmt1* and *EF1α* promoters were robustly upregulated from *tet* operators located a couple of hundred bases upstream of the transcription start sites[10]. This relaxed constraint greatly reduces the chance of affecting promoter function of a target gene in the absence of the transactivator. Increasing the number of binding sequences and/or use of stronger transactivators could help further reduce the risk by enabling upregulation from sites farther away from the transcription start site.

Our REMOTE-control system provides elegant control of the level, timing, and location of endogenous gene expression, allowing for testing the reversibility of a phenotype and the consequences of different expression levels, which are not readily achievable by current in vivo gene expression-control technologies. It is important to note that in most gene expression analyses, including ours, expression values represent the average of a population of cells among which considerable variation can be found. This heterogeneity may influence cellular decision-making processes, such as differentiation or apoptosis[58]. Though the precision of gene expression control could likely be further improved by additional genetic circuit engineering[59], the observed potency of our current system will allow useful investigation of gene function in many biological contexts. In addition, a high degree of target specificity is expected because of the complexity of the operator sequences as well as the large evolutionary distance between mammals and the originating species of the regulatory components[60]. Furthermore, transgenic mouse lines of repressors and activators can be developed and employed for any endogenous gene. For example, existing *tet* transactivator mouse models can be adapted to accomplish upregulation of a target gene in the desired mouse tissues. We recently developed a transgenic line that can drive robust tissue-specific expression of our enhanced lac repressor in multiple tissue types when combined with existing Cre lines by introducing

the *lacIGY* gene into the *Hipp11* locus[48] under the control of a Lox-STOP-Lox element (unpublished). This line would substantially facilitate the tissue-specific application of the REMOTE-control system.

Gene upregulation by the REMOTE-control system provides several advantages in comparison to current inducible transgenic approaches. It does not require generation of multiple transgenic lines to test for position effects of the insertion, as it utilizes the endogenous locus. Additionally, this approach is well-suited for upregulation of genes with strong baseline expression because it enhances expression from an already robust promoter, whereas conventional transgenic models rely on minimal viral promoters. Lastly, the tissue specificity, cell-cycle control, and splicing variants of a target gene may be retained upon upregulation by our approach, as it preserves elements of natural regulation such as innate *cis*-regulatory elements. The advent of CRISPR/Cas-mediated gene-targeting technology will greatly facilitate the application of this technology in diverse model systems.

Disclosures

PWL serves on the Scientific Advisory Boards of AnchorDx and Progenity, Inc.

Acknowledgments

References

1. Jackson-Grusby, L. et al. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genetics*. **27** (1), 31-39, (2001).
2. David, G., Turner, G. M., Yao, Y., Protopopov, A., DePinho, R. A. mSin3-associated protein, mSds3, is essential for pericentric heterochromatin formation and chromosome segregation in mammalian cells. *Genes & Development*. **17** (19), 2396-2405, (2003).
3. Sumi-Ichinose, C., Ichinose, H., Metzger, D., Chambon, P. SNF2beta-BRG1 is essential for the viability of F9 murine embryonal carcinoma cells. *Molecular and Cellular Biology*. **17** (10), 5976-5986, (1997).
4. Premsrirut, P. K. et al. A rapid and scalable system for studying gene function in mice using conditional RNA interference. *Cell*. **145** (1), 145-158, (2011).
5. Qiu, S., Adema, C. M., Lane, T. A computational study of off-target effects of RNA interference. *Nucleic Acids Research*. **33** (6), 1834-1847, (2005).
6. Gilbert, L. A. et al. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell*. **159** (3), 647-661, (2014).
7. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*. **517** (7536), 583-588, (2015).
8. Peng, H., Ivanov, A. V., Oh, H. J., Lau, Y. F., Rauscher, F. J., 3rd. Epigenetic gene silencing by the SRY protein is mediated by a KRAB-O protein that recruits the KAP1 co-repressor machinery. *The Journal of Biological Chemistry*. **284** (51), 35670-35680, (2009).
9. Groner, A. C. et al. KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading. *PLOS Genetics*. **6** (3), e1000869, (2010).
10. Lee, K. H., Oghamian, S., Park, J. A., Kang, L., Laird, P. W. The REMOTE-control system: a system for reversible and tunable control of endogenous gene expression in mice. *Nucleic Acids Research*. **45** (21), 12256-12269, (2017).
11. Gossen, M., Bonin, A. L., Bujard, H. Control of gene activity in higher eukaryotic cells by prokaryotic regulatory elements. *Trends in Biochemical Sciences*. **18** (12), 471-475, (1993).
12. Hu, M. C., Davidson, N. The inducible lac operator-repressor system is functional in mammalian cells. *Cell*. **48** (4), 555-566, (1987).
13. Cronin, C. A., Gluba, W., Scrable, H. The lac operator-repressor system is functional in the mouse. *Genes & Development*. **15** (12), 1506-1517, (2001).
14. Brand, A. H., Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*. **118** (2), 401-415, (1993).
15. Council, N. R. *Guide for the Care and Use of Laboratory Animals: Eighth Edition.* The National Academies Press, (2011).
16. Casper, J. et al. The UCSC Genome Browser database: 2018 update. *Nucleic Acids Research*. **46** (D1), D762-d769, (2018).
17. Church, D. M. et al. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLOS Biology*. **7** (5), e1000112, (2009).
18. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*. **107** (50), 21931-21936, (2010).
19. Heintzman, N. D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*. **459** (7243), 108-112, (2009).
20. Heintzman, N. D. et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics*. **39** (3), 311-318, (2007).
21. Rada-Iglesias, A. et al. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. **470** (7333), 279-283, (2011).
22. Handoko, L. et al. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genetics*. **43** (7), 630-638, (2011).
23. Splinter, E. et al. CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes & Development*. **20** (17), 2349-2354, (2006).
24. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature.* **489** (7414), 57-74, (2012).

25. Rosenbloom, K. R. et al. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research*. **41** (Database issue), D56-63, (2013).

26. Murray, J. I., Voelker, R. B., Henscheid, K. L., Warf, M. B., & Berglund, J. A. Identification of motifs that function in the splicing of non-canonical introns. *Genome Biology*. **9** (6), R97, (2008).

27. Taggart, A. J. et al. Large-scale analysis of branchpoint usage across species and cell lines. *Genome Research*. **27** (4), 639-649, (2017).

28. Corvelo, A., Hallegger, M., Smith, C. W., Eyras, E. Genome-wide association between branch point properties and alternative splicing. *PLOS Computational Biology*. **6** (11), e1001016, (2010).

29. Haeussler, M. et al. Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biology*. **17** (1), 148, (2016).

30. Grainger, S. et al. CRISPR Guide RNA Validation In Vitro. *Zebrafish*. **14** (4), 383-386, (2017).

31. Quadros, R. M., Harms, D. W., Ohtsuka, M., Gurumurthy, C. B. Insertion of sequences at the original provirus integration site of mouse ROSA26 locus using the CRISPR/Cas9 system. *FEBS Open Bio*. **5** 191-197, (2015).

32. Harms, D. W. et al. Mouse Genome Editing Using the CRISPR/Cas System. *Current Protocols in Human Genetics*. **83** 15.17.11-27, (2014).

33. Miura, H., Quadros, R. M., Gurumurthy, C. B., Ohtsuka, M. Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nature Protocols*. **13** (1), 195-215, (2018).

34. Laird, P. W. et al. Simplified mammalian DNA isolation procedure. *Nucleic Acids Research*. **19** (15), 4293, (1991).

35. Kosicki, M., Tomberg, K., Bradley, A. Repair of double-strand breaks induced by CRISPR-Cas9 leads to large deletions and complex rearrangements. *Nature Biotechnology*. **36** (8), 765-771, (2018).

36. Akcakaya, P. et al. In vivo CRISPR editing with no detectable genome-wide off-target mutations. *Nature*. **561** (7723), 416-419, (2018).

37. Lazzarotto, C. R. et al. Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nature Protocols*. **13** (11), 2615-2642, (2018).

38. Ohtsuka, M. et al. Improvement of pronuclear injection-based targeted transgenesis (PITT) by iCre mRNA-mediated site-specific recombination. *Transgenic Research*. **22** (4), 873-875, (2013).

39. Ohtsuka, M. et al. Pronuclear injection-based mouse targeted transgenesis for reproducible and highly efficient transgene expression. *Nucleic Acids Research*. **38** (22), e198, (2010).

40. Lee, G., Saito, I. Role of nucleotide sequences of loxP spacer region in Cre-mediated recombination. *Gene*. **216** (1), 55-65, (1998).

41. Thomson, J. G., Rucker, E. B., 3rd, Piedrahita, J. A. Mutational analysis of loxP sites for efficient Cre-mediated insertion into genomic DNA. *Genesis*. **36** (3), 162-167, (2003).

42. Cho, A., Haruyama, N., Kulkarni, A. B. Generation of Transgenic Mice. *Current Protocols in Cell Biology*. Chapter Unit-19.11, (2009).

43. Pu, X. A., Young, A. P., Kubisch, H. M. Production of Transgenic Mice by Pronuclear Microinjection. *Methods in Molecular Biology*. **1874** 17-41, (2019).

44. Murgha, Y. et al. Combined in vitro transcription and reverse transcription to amplify and label complex synthetic oligonucleotide probe libraries. *Biotechniques*. **58** (6), 301-307, (2015).

45. Quadros, R. M. et al. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biology*. **18** (1), 92, (2017).

46. Chen, X., Wu, J. M., Hornischer, K., Kel, A., Wingender, E. TiProD: the Tissue-specific Promoter Database. *Nucleic Acids Research*. **34** (Database issue), D104-107, (2006).

47. Haruyama, N., Cho, A., Kulkarni, A. B. Overview: Engineering transgenic constructs and mice. *Current Protocols in Cell Biology*. Chapter Unit-19.10, (2009).

48. Tasic, B. et al. Site-specific integrase-mediated transgenesis in mice via pronuclear injection. *Proceedings of the National Academy of Sciences of the United States of America*. **108** (19), 7902-7907, (2011).

49. JoVE Science Education Database. *Lab Animal Research*. Fundamentals of Breeding and Weaning. *Journal of Visualized Experiments*. (2018).

50. Wyborski, D. L., DuCoeur, L. C., Short, J. M. Parameters affecting the use of the lac repressor system in eukaryotic cells and transgenic animals. *Environmental and Molecular Mutagenesis*. **28** (4), 447-458, (1996).

51. Wyborski, D. L., Short, J. M. Analysis of inducers of the E.coli lac repressor system in mammalian cells and whole animals. *Nucleic Acids Research*. **19** (17), 4647-4653, (1991).

52. Traykova-Brauch, M. et al. An efficient and versatile system for acute and chronic modulation of renal tubular function in transgenic mice. *Nature Medicine*. **14** (9), 979-984, (2008).

53. Michel, G., Mosser, J., Fauran, F. Serum kinetics of doxycycline polyphosphate in dogs. *European Journal of Drug Metabolism and Pharmacokinetics*. **4** (1), 43-48, (1979).

54. Bertocchi, I. et al. Regulatory functions of limbic Y1 receptors in body weight and anxiety uncovered by conditional knockout and maternal care. *Proceedings of the National Academy of Sciences of the United States of America*. **108** (48), 19395-19400, (2011).

55. Plageman, T. F., Jr., Lang, R. A. Generation of an Rx-tTA: TetOp-Cre knock-in mouse line for doxycycline regulated Cre activity in the Rx expression domain. *PLOS One*. **7** (11), e50426, (2012).

56. Mollova, M. et al. Cardiomyocyte proliferation contributes to heart growth in young humans. *Proceedings of the National Academy of Sciences of the United States of America*. **110** (4), 1446-1451, (2013).

57. Ptashne, M. Principles of a switch. *Nature Chemical Biology*. **7** (8), 484-487, (2011).

58. Balazsi, G., van Oudenaarden, A., Collins, J. J. Cellular decision making and biological noise: from microbes to mammals. *Cell*. **144** (6), 910-925, (2011).

59. Nevozhay, D., Zal, T., Balazsi, G. Transferring a synthetic gene circuit from yeast to mammalian cells. *Nature Communications*. **4** 1451, (2013).

60. Labow, M. A., Baim, S. B., Shenk, T., Levine, A. J. Conversion of the lac repressor into an allosterically regulated transcriptional activator for mammalian cells. *Molecular and Cellular Biology*. **10** (7), 3343-3356, (1990).