

## Creating conditional dual fluorescence labelled transgenic animals for studying function of small non-coding RNAs

Kun Yang, Yun Gao, Mingfu Yang, Zuoshang Xu & Qian Chen

To cite this article: Kun Yang, Yun Gao, Mingfu Yang, Zuoshang Xu & Qian Chen (2016): Creating conditional dual fluorescence labelled transgenic animals for studying function of small non-coding RNAs, *Connective Tissue Research*, DOI: [10.1080/03008207.2016.1247834](https://doi.org/10.1080/03008207.2016.1247834)

To link to this article: <http://dx.doi.org/10.1080/03008207.2016.1247834>



Accepted author version posted online: 20 Oct 2016.



Submit your article to this journal [↗](#)



Article views: 3



View related articles [↗](#)



View Crossmark data [↗](#)

# Creating conditional dual fluorescence labelled transgenic animals for studying function of small non-coding RNAs

**Short title:** Small non-coding RNA transgenic mice

**Kun Yang<sup>1</sup>, Yun Gao<sup>1</sup>, Mingfu Yang<sup>1</sup>, Zuoshang Xu<sup>2</sup>, Qian Chen<sup>1</sup>**

<sup>1</sup>Department of Orthopedics, Warren Alpert Medical School of Brown University and Rhode Island Hospital, Providence, RI, USA. <sup>2</sup> Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA.

Correspondence to: Qian Chen, Department of Orthopedics, Warren Alpert Medical School of Brown University and Rhode Island Hospital, 1 Hoppin Street, Providence, RI, 02903. Email: qian\_chen@brown.edu

## **Abstract**

Because the function of most non-coding (nc) RNAs is unknown, Cre-lox transgenic mice are useful tools to determine their functions in a tissue or developmental stage-specific manner. However, the technology faces challenges because expression of ncRNA-transgene lacks protein product. No antibody or peptide-tag can be used to trace ncRNA expression in mouse tissues in real time. Furthermore, transgene integration at different locus or orientations in the genome may result in recombination of genomic fragments in the Cre-lox system. Establishing a reliable method that can be used to determine the precise copy number and orientation of the transgene is critical to the field. We developed a fast and straightforward method to determine ncRNA-transgene copy number, orientation, and insertion site in the genome. Furthermore, upon tissue-specific expression of ncRNA, a Cre-loxP mediated dual-fluorescence expression system facilitates fluorescence signal switching from green to red, which enables real time monitoring of ncRNA expression by fluorescence signals. As proof of concept, we demonstrate that after

microRNA-Flox mice crossed with Col2a1-Cre mice, microRNA transgene expression could be detected successfully by red fluorescence signals in various cartilaginous tissues. This method of creating small ncRNA transgenic mice facilitates both tissue specific ncRNA expression and real-time visualization of its expression. It is particularly suitable for *in vivo* studies of the functional roles and lineage tracing of small ncRNA.

**Keywords:** non-coding RNA, transgenic mice, microRNA, transgene copy number, droplet digital PCR, dual fluorescence switching system

### **Article History**

*Received 16 September 2016*

*Revised 30 September 2016*

*Accepted 8 October 2016*

*Published online xxxx*

## **Introduction**

In addition to messenger RNA that encodes proteins, non-coding RNAs, including microRNA (miRNA) and long non-coding RNA (lncRNA) have been shown to be involved in gene regulation at post-transcriptional and translational levels. For example, microRNAs, about 22 nucleotide long, are among the largest known classes of non-coding RNAs involved in gene silencing with more than 1,000 predicted in human (1, 2). MicroRNAs are typically encoded in

intron, and transcribed by RNA Polymerase II (3, 4). They are further processed by Drosha and Dicer complexes into mature miRNAs, which guide the binding of RNA-induced silencing complex (RISC) to the 3' UTR of target genes typically through imperfectly base pairs to induce translational repression. It is predicted that every miRNA has hundreds of target genes due to the imperfect base pairing (5, 6). About 50% of the human genes are targeted by miRNAs, however, the roles of most non-coding RNAs are unknown (5, 6).

MicroRNAs play an important role in development as shown by previous *in vivo* studies. Complete knockout of Dicer, which prevents the processing of all miRNAs, causes embryonic lethality in mice (7) and zebra fish (8). Conditional knockout of Dicer in tissue specific cell lineages revealed essential roles of miRNAs in angiogenesis (9, 10), hair follicle growth and skin development (11-13), retinal development (14, 15), musculoskeletal system development (16, 17), cortical neurogenesis (18-20), immune system development (21-23) and fertility cell development (24, 25). Conditional loss of Dicer function in specific developmental lineages provides useful information for revealing the role of miRNAs. However, tissue-specific genetic manipulation of each individual miRNA is necessary for understanding its specific function *in vivo*. Studies on gain-of-function and loss-of-function mutation of individual miRNA or miRNA cluster have been performed. Among them, most of the studies were done with transgenic (gain-of-function mutation) mice model and knockout (loss-of-function mutation) model, while only a few of them were done with knock-in (gain-of-function mutation at specific genomic location) model (26).

Creating transgenic mice is a commonly used method in animal studies because of the relatively shorter period to obtain transgenic animals in comparison to knock-in method. Furthermore, robust pronuclear microinjection technique has been optimized during the last two decades (27-30). However, this procedure relies on the random integration of transgene into mouse genome. It has been shown that in this process, transgene frequently integrates as multiple copies in tandem with varied copy numbers among transgenic lines (31). An accurate and precise quantification of transgene copy numbers, which can influence transgene expression levels, would improve the production of transgenic animals and help understand phenotype outcomes. Furthermore, when transgenes integrate as multiple copies in tandem, they can have different orientations. When loxP cassettes are oriented in the same direction, the DNA sequences flanked by them will be cut off. However, when they are oriented in opposite directions, the DNA flanked by them will be inverted. Transgenes with opposite directions of loxP can lead to chromosome loss during Cre-mediated recombination, which may cause embryo lethality or other severe phenotypes (32, 33). Although the same orientation of tandem repeats insertion is the dominant mode while opposite orientation of tandem insertions is rare (34), transgene insertion orientation needs to be determined.

Finally, studies of a non-coding RNA can be challenging because it encodes no protein products, therefore antibody-based detection methods such as western blot or immunohistochemistry cannot be used to track transgene expression in cells or tissues. Developing an easy-to-use expression monitoring system for a non-coding RNA transgene is important to locate and investigate non-coding RNA expression and its potential targets in transgenic animals. In this study, we demonstrated that a Cre-lox enabled non-coding RNA expression system with dual-

fluorescence switching facilitates both tissue specific expression of non-coding RNA and monitoring of transgene expression with fluorescence signals.

## **Material and Methods:**

### **Cloning of transgenic cDNA construct**

The pCAG G/R/M plasmid was used to clone non-coding RNA into the expression vector (Fig. 1a). The vector is pCAG-GFP/RFP-miRint-KX plasmid that contains non-coding RNA insertion sites (KpnI and XhoI) in the intron of RFP (35). In this study, the murine pre-miRNA-365-1 from chromosome 16 (GenBank acc. NC\_000082.6; bases 13453840-13453926) was cloned into the transgenic construct (36). The 111 bp DNA fragment containing mmu-miR-365-1 precursor sequence was amplified with primers PremiR365-1 fwd, PremiR365-1 rev1 and PremiR365-1 rev2. The precursor sequence was inserted into KpnI and XhoI sites. Construct was sequenced to make sure no site mutations in miR-365 precursor clone.

### **Cell culture and transfection**

Mouse NF1 were maintained in DMEM supplemented with 10% FBS (Gibco). Cells were seeded in 6 well plates the day before transfection at 80% confluence with  $2 \times 10^5$  cells in each well. Transfection was carried out with Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions 16h after seeding the cells.

## **Generation of microRNA flox transgenic mice**

Transgenic construct pCAG G/R/M-miR365 (G/R-365), as showed in Figure 1A, was linearized by digesting with ScaI and DrdI. MiR-365 flox transgenic founders were generated at Brown University Mouse Transgenic and Gene Targeting Facility by pronuclear injection into mice embryos of FVB strain with standard protocols. Several founders within forty-three pronuclear injected mice embryos (F0 generation) were identified by PCR genotyping as miR-365 transgene positive.

C57/BL6 mice from Jackson laboratories were used to back cross with the founders for 8 generations. Three independent lines were used in further characterization and analysis. All mice were housed, handled, and euthanized in accordance with federal and institutional guidelines. Animal protocols approved by Lifespan IACUC animal studies committee for this project are CMTT#0128-10 and #0081-13.

### **PCR genotyping**

For genomic DNA isolation, tail snips were treated with lysis buffer plus proteinase K at 56 °C water bath for 2 hours. Three sets of primers were designed to characterize miR-365 transgenic mice in details. They are designed across the whole construct to make sure the transgene in founders is not truncated. The first set of primers with the product shown as the red bar in Figure 1A at CMV Enhancer region are CMVE fwd and CMVE rev. The second set of primers with the product showed as the red bar in Figure 1a at RFP encoding region are RFP fwd and RFP rev. The third set of primers with the product showed as the red bar in Figure 1a at the conjunction of

RFP and miR-365 region are MiR365fl fwd and MiR365fl rev. After pronuclear injection, 43 mice were born after being injected. They were genotyped with primer set 2 and 7 founders were identified as transgene positive. After transgene copy number screening and orientation determination, three lines were selected. They were further investigated for the integrity of transgene by genotyping with all three sets of primers, which located at the beginning, middle, and tail of the transgene, separately.

### **Fluorescence signal examination**

Fluorescence images for cultured cells were captured with Nikon Eclipse T100 microscope equipped with FITC-HYQ filter for green fluorescence and TRITC-HYQ filter for red fluorescence. Tissues from transgenic animals were fixed with 4% paraformaldehyde in PBS solution followed by impregnating in 30% sucrose overnight. Tissues were embedded in OCT compound and cut by cryostat (Reichert-Jung Cryocut 1800, USA) along sagittal plane of the knee joint at 30 $\mu$ m thickness. Slides were visualized by using a confocal fluorescence microscope (Nikon C1Si Confocal Inverted Microscope).

### **Determining mature miRNA expression level by Real Time PCR analysis**

To detect miR-365 expression level by real time PCR, RNA was isolated from tissue or cells using miRNeasy kit (Qiagen217004) according to the manufacturer's protocol. Then cDNA was synthesized using miScript II RT kit (Qiagen 218161). Quantitative real time PCR was

performed on a DNA engine 2 (Bio-rad) using SYBR green master mix (Invitrogen 4367659). MiRNA expression levels were analyzed using the miRNA primer assays of miR-365 (Qiagen MS00032760) with RNU6 primer assay (Qiagen MS00033740) as the control according to manufacturer's instructions.

## **Determining transgene copy numbers in founders (F0) by Real Time PCR analysis**

The principle for transgene copy number determination with real-time PCR analysis is estimating unknown animal samples by comparing them with a standard curve fitted with known plasmid DNAs. Briefly, the standard curve with constructed transgene plasmid (copy number is known by DNA quantity measurement) is firstly calibrated and subsequently the transgene quantity by a pair of primers amplifying a 150bp fragment of RFP gene is interpolated into that plasmid standard curve to obtain copy numbers of samples. The mass of mouse DNA is roughly 3.3pg per haploid copy of genome. The transgene construct with about 8.7k base pair of nucleic acids equals to  $0.95 \times 10^{-17}$ g. Transgene construct was serially diluted from  $7.32 \times 10^{-2}$ pg which equals to 64 copies by 2 to  $1.14 \times 10^{-3}$ pg which equals to 1 copy of transgene. 396pg genomic DNA, which equals to 120 copies of haploid genome, isolated from wild type mice was added in previous seven serial dilutions as well as the eighth tube with no plasmid as negative control. The same amount of 396pg genomic DNA from each transgenic mouse was tested in real time PCR for 150bp RFP fragment with another 195bp Fabpi (intestinal fatty acid binding protein) served as internal control.  $-\Delta C_t$  values, that is, the difference between  $C_t$  value of the internal control and  $C_t$  value of RFP, were plotted versus  $\log_2$  transformed copy numbers and a simple

linear regression was applied to build the standard curve(Fig. 2A).To determine transgene copy number by quantitative PCR, genomic DNA was cleaned by DNeasy blood & tissue kit (Qiagen 69504). The primers used for quantitative PCR reaction are: Forward primer RFP-F and reverse primer RFP-R for RFP gene, forward primer int-F1 and reverse primer int-R1 for the internal reference gene Fabpi.

To improve the accuracy of real-time PCR estimation for copy numbers, two dilutions from each animal genomic DNA samples were tested instead of one in the previous method. A multiple linear regression model was fitted with both the Ct values from plasmid dilutions and the animal sample dilutions (Fig. 2C). It is for getting better estimation of the standard curve slope by adjusting the slope with more degree of freedom. In other words, the  $\Delta C_t$  values for all the founders could be used to train the standard curve and get the empirical slope from this regression and then estimate copy numbers using this slope.

## **Determining transgene copy numbers with Droplet digital PCR (ddPCR) analysis**

Genomic DNAs were cleaned with DNeasy blood & tissue kit (Qiagen 69504) to remove any detergent in the lysis buffer which will prevent droplet partition. Genomic DNA was sheared into 3kb by sonication with Qsonicsonicator (Fisher Scientific). Prior to sonication, the mouse DNA samples were diluted to 0.01  $\mu\text{g}/\mu\text{l}$  and a final volume of 100  $\mu\text{l}$ . The sonication was set on low power and sonication was performed with 30 sec ON, 30 sec OFF cycles. To achieve 3 kb fragment, 4 to 5 cycles were desired when samples were kept on ice during sonication. The DNA

fragments were examined with agarose gel electrophoresis prior to ddPCR. The probe used for copy number detection is EGFP probe labeled with FAM (Life technology 4400291) with the internal control of mouse Tert labeled with VIC (Life technology 4458368). The PCR reaction was prepared in 25  $\mu$ l final volume, using 12.5  $\mu$ l TaqMan Genotyping Master Mix (Life Technologies), 2.5  $\mu$ l Droplet Stabilizer (RainDance Technologies, Lexington, MA), 2.5  $\mu$ l Assay Mix (10 $\times$ ) and 7.5  $\mu$ l cDNA template in water. Droplets containing PCR reaction components were generated using a hydrodynamic flow-focusing microfluidic chip (RainDrop Source chip; RainDance Technologies), deposited into PCR tubes as 5  $\mu$ l aqueous droplets suspended in inert fluorinated oil (REB Carrier Oil; RainDance Technologies), and subjected to PCR amplification in a thermal cycler (LifeECO; BIOER). PCR cycling parameters: 95  $^{\circ}$ C for 10 minutes, then 44 cycles of 95  $^{\circ}$ C for 15 seconds and 60  $^{\circ}$ C for 1 minute. After PCR completion, the emulsion was injected into a second microfluidic chip (RainDrop Sense chip; RainDance Technologies) for fluorescence measurement. Following 488 nm excitation, droplet fluorescence was detected through filters (FAM and VIC centered emission) with photomultiplier tubes recording fluorescence intensity (“height”) and duration (“width”). Spectral crosstalk-corrected data from each sample or control was converted to a two-dimensional (FAM and VIC intensity) histogram (electronic and photonic noise was removed; droplet data was width filtered and normalized to background fluorescence). Custom software (RainDrop Analyst; RainDance Technologies) was used to define graphical areas or “gates” associated with EGFP or Tert and count the number of droplets within each gate.

## **Determination of transgene insertion orientations by PCR**

To check the junction orientation of multiple copies of transgene, primers were designed for three possible junctions as head-to-tail, head-to-head and tail-to-tail. One set of primers were used for head-to-tail junction, two sets were used for head-to-head junction and three sets were used for tail-to-tail junctions. Head-to-tail primers are HTF1 and HTR1. Head-to-head primers are: HHF1, HHR1 and HHR2. Tail-to-tail primers are: TTF1, TTF2, TTR1, TTR2 and TTR3.

## **Determination of transgene insertion location by inverse PCR**

Transgenic animals were bred with wild type C57/BL6 mice for 8 generations to allow chromosome segregation and switch mouse genetic background into C57/BL6. Tail snip isolated genomic DNA was purified with DNeasy Blood & Tissue kit (Qiagen Cat. No. 69504). 2 $\mu$ g DNA template was digested with AgeI, AflIII or PmeI (NEB, MA), separately. Digested DNA was purified by standard phenol:chloroform procedure and then dissolved in TE at a concentration of 100 $\mu$ g/ml. Self-ligation reactions for each digested DNA samples were set up with the final concentration of cleaved template DNA at ranging from 0.1 to 1  $\mu$ g/ml. Ligation reactions were carried out for 12-16 hours at 16°C. The ligated DNA was purified with standard phenol:chloroform procedure and then dissolved in 10 mM Tris (pH 7.6) at a concentration of 100  $\mu$ g/ml. The ligation product was then digested with EcoRV for linearization. Digested Products were then amplified with nested PCR with 1st set of primers: forward primer IPCR F1 and reverse primer IPCR R1; and then 2nd set of primers: forward primer IPCR F2 and reverse primer IPCR R2. PCR amplified genomic DNA were then sequenced with both IPCR F2 and

IPCR R2. Sequencing results were further confirmed with PCR amplification using forward primer designed base on genomic DNA sequencing result and reverse primer on the transgene construct: forward primer Ch4F1 and reverse primer IPCR R2. The PCR product is about 200bp.

## **Statistics Analysis**

All statistical testing was performed using unpaired two-tailed Student's T-test. P-values  $\leq 0.05$  were considered statistically significant. Error bars represent standard error of the mean (s.e.m.).

## **Results:**

### **ncRNA expression visualized by a switch of fluorescence from GFP to RFP**

The non-coding RNA expression vector contains a CAG promoter and a CMV enhancer that drives the expression of green fluorescence protein (GFP) (Fig. 1A). In the presence of Cre recombinase, the cDNA sequence encoding GFP and Poly-A, which is flanked by LoxP cassettes, is spliced out. As a result, the promoter and enhancer drive the expression of both RFP and a non-coding RNA that is inserted into the intron of RFP (Fig. 1A).

To determine whether the switch of GFP to RFP corresponds to the turning on of noncoding RNA expression, we inserted microRNA-365 sequence into the intron of RFP in the F1 vector (Fig. 1A). Before pronuclear injection to generate transgenic mice, the transgene construct was tested in NF1 mouse cell lines in vitro (Fig. 1B, C). Transfection of miR-365 F1 vector alone

resulted in green fluorescence in the transfected cells (Fig. 1C) without expression of miR-365 in transfected cells (Fig. 1B). Co-transfection of CMV-Cre results in 25-fold increase of miR-365 (Fig. 1B), reduction of green fluorescent cells, and appearance of red fluorescent cells (Fig. 1C). When co-transfection ratio was changed to 1:3(miR-365 fl : CMV-Cre) without altering total amount of transfected plasmid, more red cells and less green cells were observed (Fig. 1C). It indicated a more complete conversion due to the increase of Cre recombinase activity.

### **Screening transgenic lines by genotyping and transgene copy number determination**

Among forty three pronuclear injected animals (F0 generation), seven miR-365 flox founders were identified as miR-365 transgene positive by PCR genotyping. They were screened for transgene copy numbers by real-time PCR analysis. A standard curve was fitted with transgene construct as known control with serial dilutions (Fig. 2A). Transgene copy numbers of seven founders were estimated by calculation based on the standard curve. Results showed that copy numbers of seven founders varied from 1 to 38 (Fig. 2B). The founders with low copy number T28-1C (2.4±0.8), T29-1B (4.2±2.2) and T29-3H (1.4±1.6) were selected for further characterization, since high copy numbers have been associated with non-specific effects (1).

The accuracy of the copy number determination relies on the estimation of standard curve slope by real-time PCR. To achieve better statistical outcome, a multiple linear regression model was fitted with both the Ct values from plasmid dilutions and the animal sample dilutions (Fig. 2C). With this model, the estimated copy numbers have a smaller deviation, which gives a better

estimation of the true values with smaller confidence intervals. Multiple linear regression analysis indicated that T28-1C (2.8+0.4), T29-1B (3.8+0.6) and T29-3H (2.2+1) were still the best candidates.

Finally, we developed a fast and straightforward method to determine ncRNA-transgene copy number with Raindance droplet digital polymerase chain reaction as described in details in Methods section. Absolute values of copy numbers obtained from ddPCR analysis are T28-1C with two copies, T29-1B with five copies and T29-3H with two copies of transgenes (Fig. 1D).

### **Determination of transgene integrity and insertion orientation**

To make sure the entire transgene was successfully inserted into the mouse genome, the integrity of the transgene was determined by three sets of primers (red bars in Figure 1A) in each transgenic line. The PCR results showed that F1#17 from founder T28-1C, #21 from founder T29-1B and #46 from founder T29-3H all contained the predicted fragments in the different areas of the transgene indicated by positive products from PCR using three sets of genotyping primers (Fig. 3A).

When transgenes integrate as multiple copies in tandem, they can have different orientations (1). The tandem head-to-tail (HT) orientation with loxP in the same direction is often the dominant mode while head-to-head (HH) or tail-to-tail (TT) with loxP in opposite directions are rare cases when the copy number is low (Fig. 3D). When loxP cassettes are oriented in the same direction, the DNA sequences (B and C) flanked by them will be cut off, and A and D will be connected as intended (Fig. 3B). When they are oriented in opposite directions, the DNA flanked by them (B

and C) may be inverted (C and B) (Fig. 3C). Thus, transgenes with opposite directions of loxP can lead to unintended chromosome rearrangement or loss during cre-mediated recombination, which may cause embryo lethality or other severe phenotypes (2, 3). PCR analysis was performed using primer sets designed to detect transgene orientations (Fig. 3D). The results showed that the low copy number F1#17, #21 and #46 only contains head-to-tail oriented transgenes, while the high copy number founder T29-1C contains transgenes with all three orientations (Fig. 3E).

### **Identification of the transgene insertion site in the genome**

Since one of the transgenic lines T28-1C (Fig. 2), contains only two transgene copies with head-to-tail orientation, it indicates that there is only one transgene insertion site in mouse chromosomes. To identify the transgene insertion site, we performed inverse PCR with the procedures illustrated in Fig. 4A. The purpose of this analysis is to identify the genomic DNA sequences surrounding the transgene. Genomic DNA from miR-365 flox transgenic mice (T28-1C line) were digested with restriction enzyme and circularized by re-ligation. The ligation products were then cut within the transgene sequence by EcoR V. Finally, the linearized genome sequences were amplified with nested PCR, which were sequenced to identify the DNA sequences surrounding the transgene in mouse chromosomes. DNA sequences in both chromosome 11 and 4 were identified. While chromosome 4 does not contain any endogenous miR-365 gene, chromosome 11 contains the endogenous miR-365-2 gene. Furthermore, the identified DNA sequences were close to the genomic site of miR-365-2 gene in chromosome 11. Therefore, we concluded that the insertion site of miR-365 transgene is in chromosome 4 (Fig.

4B). The genomic location of the transgene was further confirmed by PCR reactions using the primers containing surrounding genomic DNA sequences (Fig. 4C). DNA sequence analysis indicated that the transgene was inserted into an intergenic region on chromosome 4 (Fig. 4B).

## **Fluorescence signals facilitate real time monitoring transgene expression**

Up-regulation of miR-365 in cartilaginous tissues was achieved by crossing miR-365 fl +/- mice with Col2a1-Cre +/- mice. It resulted in more than six-fold over-expression of miR-365 levels in cartilage from miR-365 fl +/-; Col2a1-Cre +/- mice in comparison to control mice (Fig. 5B). Cartilaginous tissues exhibit fluorescence switch from green fluorescence to red fluorescence, shows here with distal femur, proximal femoral head, lower jaw and spinal column as examples (Fig. 5A). The red fluorescence was strongest in the chondroprogenitor cell populations in the superficial zone of articular cartilage, which indicates the highest Col2a1-Cre activity (4, 5). Thus, this fluorescence switching system enables the real-time monitoring of ncRNA transgene expression and lineage tracing in vivo. Furthermore, the cartilage-specific over-expression of miR-365 resulted in a significant reduction of femur length in miR-365 fl +/-; Col2a1-Cre +/- mice (Fig. 5C), suggesting miR-365 is involved in regulation of endochondral bone formation.

## Discussion:

Creating transgenic animals has been extremely useful for understanding the function of a gene in vivo. However, deleting a non-coding RNA gene often results in no or very subtle phenotypes, probably due to the overlapping of multiple targets by other non-coding RNAs (6). On the other hand, over-expression of a non-coding RNA, mimicking its induction under certain physiological or pathological conditions, may yield a clue to its function by affecting its multiple target genes. Since most ncRNAs are expressed in multiple tissues, it is important to create tissue-specific transgenic mice to discern its function in a particular tissue.

Cre-lox system is a great way to create a tissue or developmental stage specific transgenic mouse. However, it has been shown that during pronuclear microinjection, transgenes frequently integrate as multiple copies in tandem at random locations of the genome (7). Transgenes of high copy numbers, especially at different insertion sites or opposite orientations, may result in chromosome defect upon Cre-lox recombination (1-3). Therefore, it is essential to have a fast and efficient way to screen the transgenic lines by determining the transgene copy number, orientation, and insertion site.

We have developed such a method in this study. The approach was applied for CNS knockdown (37). But it did not work due to unintended effects from inverted tandem integrations. We demonstrate that this problem can be solved by screening for lines with low transgene copy number and ensure that the transgenes are inserted in a tandem head to tail orientation. The method consists of three steps: first, quantifying transgene copies by real-time or digital droplet

PCR; second, determining the orientation of the transgenes by PCR using orientation specific primers; and third, determining the insertion site of transgene by inverse PCR and DNA sequencing.

For quantifying transgene copies, we performed a comparison between real-time and digital droplet PCR. While real-time PCR method is a dominant method for estimating transgene copy numbers currently (33), the accuracy and detection limit needs to be improved. Its copy number estimation may be inconsistent with Southern blot analysis (34) and the detection limit of copy number is 2 (46). The transgene copy number is estimated by fitting the traditional simple linear regression standard curve, which is limited by statistical analysis. In this case, the more dilutions are performed, the more robust the standard curve will be. However, there is an empirical limit for how many times the serial dilutions can be performed in experiments. We improved the accuracy by using a multiple linear regression model fitted with serial dilutions of both plasmid DNA and DNA from animals. With the adjusted slope of the standard curve, the estimated copy numbers have a smaller deviation, which gives a better estimation of true values with smaller confidence intervals.

Another limit of real-time PCR is that a standard curve was developed using the transgene plasmid plus genomic DNA from wild type animals, while the actual samples were from transgenic mice. We demonstrated that droplet digital PCR is a superior method by enabling straightforward and precise quantification of transgene copy numbers (35, 36). The results of ddPCR gave very precise values of transgene copy numbers because it quantified the copy numbers of transgene and reference gene in the same genomic DNA content.

We also determined the orientation and insertion site of transgene. Since the transgene incorporation rate is low (only 10%-15% pronuclear microinjected embryos carry the transgene), incorporation at different genomic locations is rare (about 1% to 3% possibility) (47). In other words, multiple copies of transgene are most probably located in tandem at one chromosome location. However, it is still critically important to determine that the tandem repeats are in head-to-tail arrangement to prevent any unintended recombination using the Cre-lox system (1).

We developed a dual fluorescence switching system to monitor expression of ncRNA in tissues. Co-expression of a reporter gene with ncRNA driven by the same promoter facilitates easy observing transgene expression location and intensity in desired tissues. This overcomes the deficiency of analysis using antibody or peptide-tag due to the lack of protein product by ncRNA gene expression. This is particularly important for studying hard tissues like bone and adjacent cartilage tissues. Although in situ hybridization has been used to track RNA expression, for skeletal tissues in adult animals, it's necessary to decalcify the specimen before embedding and sectioning. This whole process is very harsh for preserving RNA content in tissue sections, which is known to degrade easily. Co-expression of red fluorescent protein with ncRNA enables direct visualization of ncRNA tissue expression patterns by red fluorescence without any requirement for treatment on frozen sections. Alternatively, the fluorescence protein markers can also be detected by immunohistochemistry on paraffin sections.

After breeding with Col2a1-Cre mice, the red fluorescence signals appeared in the cartilaginous tissues throughout the body. The red fluorescence distribution corresponds to the location and levels of miR-365 transgene expression. Furthermore, the cartilage specific expression of miR-

365 results in a reduction of femur length during skeletal development. This is consistent with its proposed role in promoting chondrocyte hypertrophy during endochondral ossification (8). MiR-365 has been shown to be expressed in different types of tissues (8-11). In particular, its expression in cartilage is up-regulated by cyclic loading, similar to miR-140, which is abundantly expressed in cartilage (42). The cartilage-specific miR-140 transgenic mice resulted in 20% over-expression of miR-140 in cartilage (12). In comparison, our Cre-lox system resulted in 6 fold over-expression of miR-365 in the transgenic mice cartilage (Fig. 5B), which is similar to its activation by mechanical loading (42). Due to the scope and focus of this manuscript, detailed phenotype analysis of miR-365fl; Col2a1-Cre transgenic mice will be presented in a separate study.

There are several advantages of using our Cre-mediated miR fluorescent marker conversion mice for lineage tracing studies. First, since a miR is turned on at the same time as the GFP to RFP switch, it is not necessary to cross miR transgenic mice with the lineage specific marker mice. Second, one can study the effect of miR in a particular cell lineage by comparing the difference between over-expression of a particular miRNA to its negative control (scrambled miRNA) mice. Third, the level of miR transgene expression is correlated with the intensity of red fluorescence since both RFP and miR are under the same promoter. It reflects the miR expression changes in a particular cell lineage over time.

In conclusion, we have developed a strategy for creating and selecting conditional dual fluorescence labeled ncRNA transgenic mice. Alternatively, gene knock-in or editing technology can also be used in generating gain-of-function mice (13). However, the genetic approach

described here is more mature, more affordable and less time consuming. With a single known transgene insertion site in the genome, the miR-365 transgenic mouse strain developed in this study is equivalent to a knock-in transgenic strain functionally. The methods described here enable rapid screening of transgenic mice for successful functional study of small ncRNAs in specific tissues or, by using tamoxifen-induction, during specific ages *in vivo*. Since intronic long ncRNAs also exist in genome, it is possible that these methods can also be used for creating lncRNA transgenic mice. This remains to be tested.

## **Acknowledgement:**

This study is supported by NIH P20GM104937 to QC and 1R21NS095635 and 1R21NS092127 to ZX. We thank Dr. Wentian Yang for critical reading of this paper and providing Col2a1-cre mouse.

## **Declaration of Conflict of Interest**

The authors report no conflict of interest

## Reference:

1. Bentwich I, Avniel A, Karov Y, Aharonov R, Gilad S, Barad O, et al. Identification of hundreds of conserved and nonconserved human microRNAs. *Nature genetics*. 2005;37(7):766-70.
2. Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, et al. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell*. 2007;129(7):1401-14.
3. Lin SL, Chang D, Wu DY, Ying SY. A novel RNA splicing-mediated gene silencing mechanism potential for genome evolution. *Biochemical and biophysical research communications*. 2003;310(3):754-60.
4. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004;116(2):281-97.
5. Plasterk RH. Micro RNAs in animal development. *Cell*. 2006;124(5):877-81.
6. Shivdasani RA. MicroRNAs: regulators of gene expression and cell differentiation. *Blood*. 2006;108(12):3646-53.
7. Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, et al. Dicer is essential for mouse development. *Nature genetics*. 2003;35(3):215-7.
8. Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, et al. MicroRNAs regulate brain morphogenesis in zebrafish. *Science*. 2005;308(5723):833-8.

9. Yang WJ, Yang DD, Na S, Sandusky GE, Zhang Q, Zhao G. Dicer is required for embryonic angiogenesis during mouse development. *The Journal of biological chemistry*. 2005;280(10):9330-5.
10. Suarez Y, Fernandez-Hernando C, Yu J, Gerber SA, Harrison KD, Pober JS, et al. Dicer-dependent endothelial microRNAs are necessary for postnatal angiogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(37):14082-7.
11. Andl T, Murchison EP, Liu F, Zhang Y, Yunta-Gonzalez M, Tobias JW, et al. The miRNA-processing enzyme dicer is essential for the morphogenesis and maintenance of hair follicles. *Current biology : CB*. 2006;16(10):1041-9.
12. Teta M, Choi YS, Okegbe T, Wong G, Tam OH, Chong MM, et al. Inducible deletion of epidermal Dicer and Drosha reveals multiple functions for miRNAs in postnatal skin. *Development*. 2012;139(8):1405-16.
13. Yi R, O'Carroll D, Pasolli HA, Zhang Z, Dietrich FS, Tarakhovsky A, et al. Morphogenesis in skin is governed by discrete sets of differentially expressed microRNAs. *Nature genetics*. 2006;38(3):356-62.
14. Georgi SA, Reh TA. Dicer is required for the transition from early to late progenitor state in the developing mouse retina. *The Journal of neuroscience : the official journal of the Society for Neuroscience*. 2010;30(11):4048-61.
15. Iida A, Shinoe T, Baba Y, Mano H, Watanabe S. Dicer plays essential roles for retinal development by regulation of survival and differentiation. *Investigative ophthalmology & visual science*. 2011;52(6):3008-17.

16. O'Rourke JR, Georges SA, Seay HR, Tapscott SJ, McManus MT, Goldhamer DJ, et al. Essential role for Dicer during skeletal muscle development. *Developmental biology*. 2007;311(2):359-68.
17. Kobayashi T, Lu J, Cobb BS, Rodda SJ, McMahon AP, Schipani E, et al. Dicer-dependent pathways regulate chondrocyte proliferation and differentiation. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(6):1949-54.
18. Nowakowski TJ, Mysiak KS, O'Leary T, Fotaki V, Pratt T, Price DJ. Loss of functional Dicer in mouse radial glia cell-autonomously prolongs cortical neurogenesis. *Developmental biology*. 2013;382(2):530-7.
19. Byon JC, Padilla SM, Papayannopoulou T. Deletion of Dicer in late erythroid cells results in impaired stress erythropoiesis in mice. *Experimental hematology*. 2014;42(10):852-6 e1.
20. Mang GM, Pradervand S, Du NH, Arpat AB, Preitner F, Wigger L, et al. A Neuron-Specific Deletion of the MicroRNA-Processing Enzyme DICER Induces Severe but Transient Obesity in Mice. *PloS one*. 2015;10(1):e0116760.
21. Oh SY, Brandal S, Kapur R, Zhu Z, Takemoto CM. Global microRNA expression is essential for murine mast cell development in vivo. *Experimental hematology*. 2014;42(10):919-23 e1.
22. Krill KT, Gurdziel K, Heaton JH, Simon DP, Hammer GD. Dicer deficiency reveals microRNAs predicted to control gene expression in the developing adrenal cortex. *Molecular endocrinology*. 2013;27(5):754-68.

23. Kuipers H, Schnorfeil FM, Fehling HJ, Bartels H, Brocker T. Dicer-dependent microRNAs control maturation, function, and maintenance of Langerhans cells in vivo. *Journal of immunology*. 2010;185(1):400-9.
24. Korhonen HM, Meikar O, Yadav RP, Papaioannou MD, Romero Y, Da Ros M, et al. Dicer is required for haploid male germ cell differentiation in mice. *PloS one*. 2011;6(9):e24821.
25. Wang H, Graham I, Hastings R, Gunewardena S, Brinkmeier ML, Conn PM, et al. Gonadotrope-specific Deletion of Dicer Results in Severely Suppressed Gonadotropins and Fertility Defects. *The Journal of biological chemistry*. 2015;290(5):2699-714.
26. Wienholds E, Plasterk RH. MicroRNA function in animal development. *FEBS letters*. 2005;579(26):5911-22.
27. Gordon JW, Scangos GA, Plotkin DJ, Barbosa JA, Ruddle FH. Genetic transformation of mouse embryos by microinjection of purified DNA. *Proceedings of the National Academy of Sciences of the United States of America*. 1980;77(12):7380-4.
28. Brinster RL, Chen HY, Trumbauer M, Senear AW, Warren R, Palmiter RD. Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell*. 1981;27(1 Pt 2):223-31.
29. Costantini F, Lacy E. Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature*. 1981;294(5836):92-4.
30. Wagner EF, Stewart TA, Mintz B. The human beta-globin gene and a functional viral thymidine kinase gene in developing mice. *Proceedings of the National Academy of Sciences of the United States of America*. 1981;78(8):5016-20.

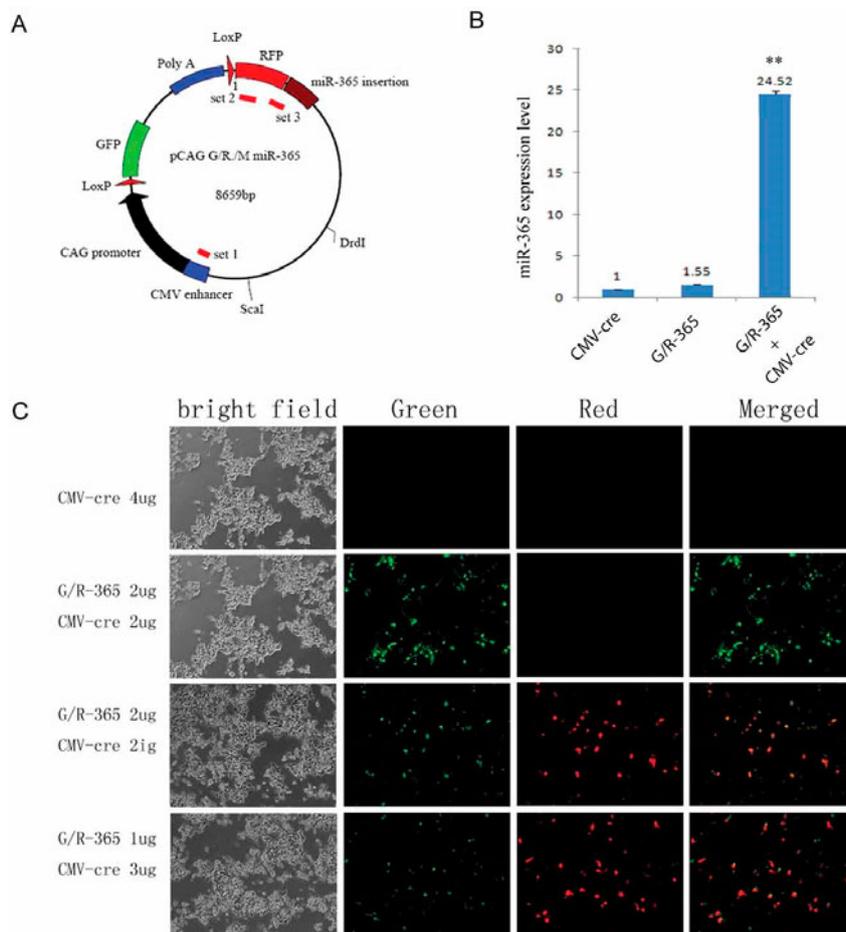
31. Niemann H, Kues WA. Transgenic livestock: premises and promises. *Animal reproduction science*. 2000;60-61:277-93.
32. Lewandoski M, Martin GR. Cre-mediated chromosome loss in mice. *Nature genetics*. 1997;17(2):223-5.
33. Gregoire D, Kmita M. Recombination between inverted loxP sites is cytotoxic for proliferating cells and provides a simple tool for conditional cell ablation. *Proceedings of the National Academy of Sciences of the United States of America*. 2008;105(38):14492-6.
34. Kopertekh L, Schulze K, Frolov A, Strack D, Broer I, Schiemann J. Cre-mediated seed-specific transgene excision in tobacco. *Plant molecular biology*. 2010;72(6):597-605.
35. Yang CX, Qiu LH, Xu ZS. Specific Gene Silencing Using RNAi in Cell Culture. *Methods in Molecular Biology*. 2011;793:457-77.
36. Griffiths-Jones S. The microRNA Registry. *Nucleic acids research*. 2004;32(Database issue):D109-11.
37. Qiu L, Rivera-Perez JA, Xu Z. A non-specific effect associated with conditional transgene expression based on Cre-loxP strategy in mice. *PloS one*. 2011;6(5):e18778.
38. Ono N, Ono W, Nagasawa T, Kronenberg HM. A subset of chondrogenic cells provides early mesenchymal progenitors in growing bones. *Nature cell biology*. 2014;16(12):1157-67.
39. Kronenberg HM. Developmental regulation of the growth plate. *Nature*. 2003;423(6937):332-6.
40. Ambros V. The functions of animal microRNAs. *Nature*. 2004;431(7006):350-5.

41. Yan BW, Zhao YF, Cao WG, Li N, Gou KM. Mechanism of random integration of foreign DNA in transgenic mice. *Transgenic Res.* 2013;22(5):983-92.
42. Guan YJ, Yang X, Wei L, Chen Q. MiR-365: a mechanosensitive microRNA stimulates chondrocyte differentiation through targeting histone deacetylase 4. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology.* 2011;25(12):4457-66.
43. Issler O, Chen A. Determining the role of microRNAs in psychiatric disorders. *Nature reviews Neuroscience.* 2015;16(4):201-12.

Accepted Manuscript

**Figure 1.** Dual fluorescence/ncRNA transgene construct and in vitro testing

(A) MiR-365 flox transgenic construct structure. Mmu-miR-365-1 was cloned into transgene construct which adapted cre/loxP system to facilitate fluorescence switching from GFP to RFP. (B) Real-time PCR analysis of miR-365 level showed that miR-365 increased for about 25 fold in miR-365 flox construct and CMV-cre co-transfected NF1 cells than single plasmid transfection. For single transfection, 4  $\mu$ g CMV-cre or miR-365 flox construct pCAG\_G/R/M\_miR-365 was transfected in six well plates with lipofectamine 2000. For double transfection, 3  $\mu$ g of CMV-cre and 1  $\mu$ g of pCAG\_G/R/M\_miR-365 was transfected. (C) Fluorescence signal showed that CMV-cre plasmid expressed no fluorescence, miR-365 flox construct pCAG\_G/R/M\_miR-365 expressed GFP and co-transfection of miR-365 flox construct and CMV-cre expressed RFP. Student t-test was used for statistics. (\*)  $P < 0.05$ ; (\*\*)  $P < 0.01$ .



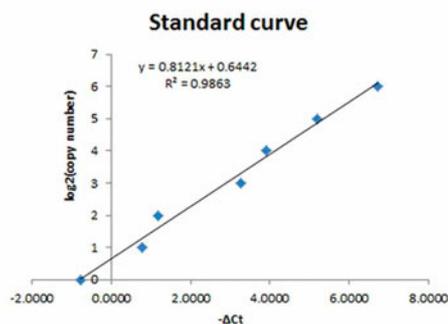
**Figure 2.** Determining transgene copy numbers in F0 generation founders

(A) Transgene plasmid was used as known control to fit simple linear regression. Serial dilutions of plasmid by adding genomic DNA from wild type mice were calculated in the table. Simple linear regression equation is right below the table. The standard curve was fitted by plotting log<sub>2</sub> transformed copy numbers versus  $-\Delta Ct$ . (B) Copy numbers of each founder were estimated using standard curve. (C) Multiple linear regression was fitted by including the slope information estimated for each tested animal. Equation is showed as well as new estimations based on multiple linear regression. (D) Droplet digital PCR was a new developed technique, which performs PCR more accurately. Copy numbers were checked by using EGFP probe for transgene copy numbers and comparing with Tert as known copy number control. Three lines, which were going to be used for breeding, were checked for copy numbers with ddPCR.

A

copy number #	1	2	4	8	16	32	64
genomic DNA (pg)	396	396	396	396	396	396	396
plasmid DNA ( $10^{-3}$ pg)	1.14	2.29	4.58	9.15	18.3	36.6	73.2

$$\log_2(\text{copy number}) = \beta_0 + \beta_{Ct} X_{Ct} + \varepsilon$$



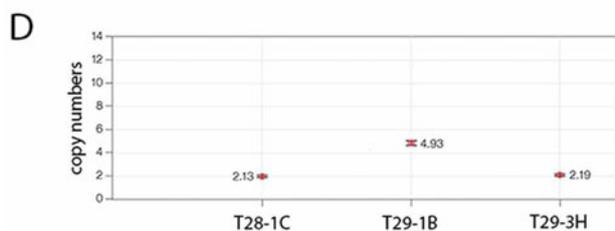
B

Founder	T28-1C	T29-1B	T29-1C	T29-1E	T29-3B	T29-3H	T29-3J
haploid	1.2 ± 0.4	2.1 ± 1.1	18.95 ± 5.3	3.72 ± 0.9	14.51 ± 3.7	0.7 ± 0.8	5.1 ± 2.3
diploid	2.4 ± 0.8	4.2 ± 2.2	38.0 ± 10.6	7.44 ± 1.8	29.02 ± 7.4	1.4 ± 1.6	10.2 ± 4.6

C

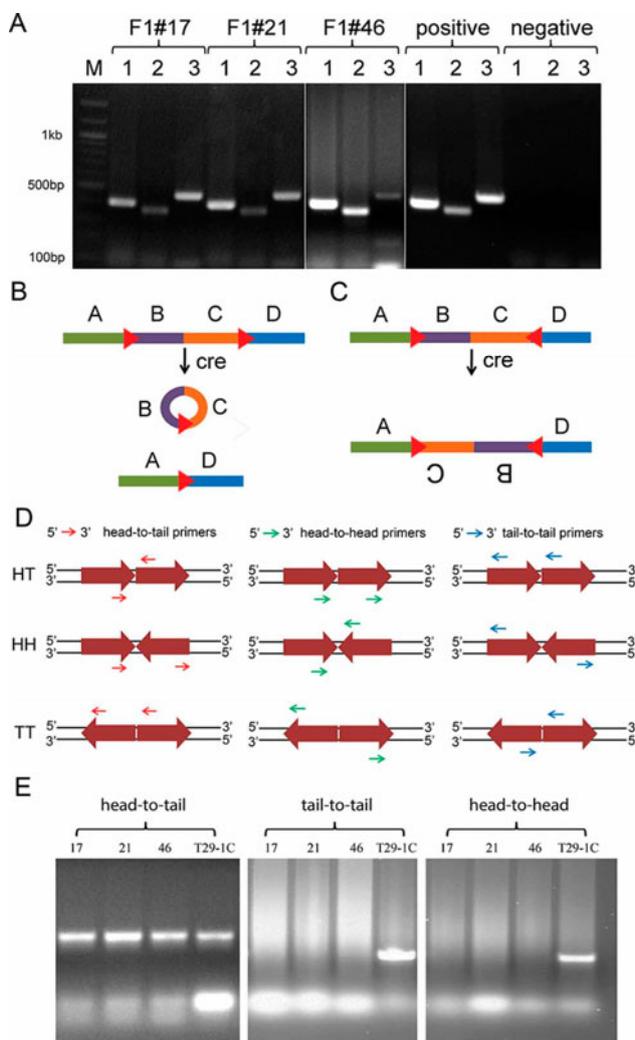
$$\log_2(\text{copy number}) = \beta_0 + \beta_1 X_{\text{plasmid}} + \beta_2 X_{\text{mice1}} + \dots + \varepsilon$$

Founder	T28-1C	T29-1B	T29-1C	T29-1E	T29-3B	T29-3H	T29-3J
haploid	1.4 ± 0.2	1.9 ± 0.3	19.55 ± 1.3	4.1 ± 0.5	16.35 ± 0.8	1.1 ± 0.5	6.4 ± 0.5
diploid	2.8 ± 0.4	3.8 ± 0.6	39.1 ± 2.6	8.2 ± 1.0	32.7 ± 1.6	2.2 ± 1.0	12.8 ± 1.0



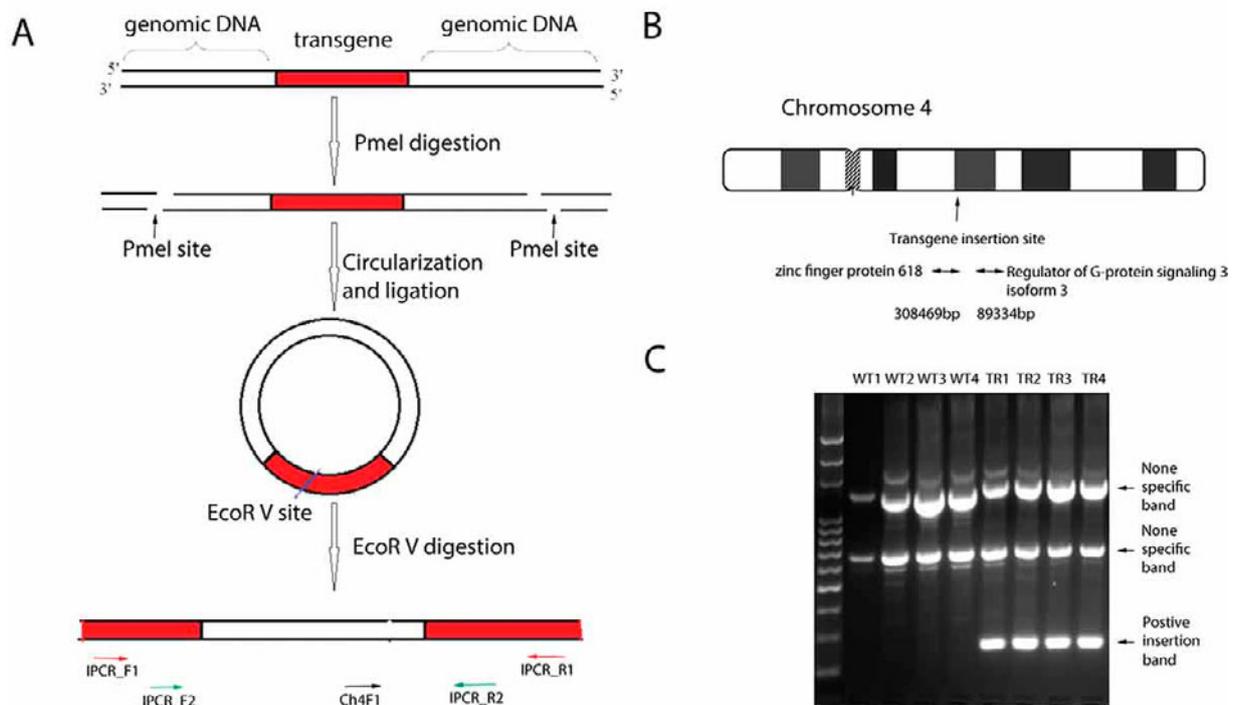
**Figure 3.** Characterization of transgene orientation

(A) Three sets of primers were used to characterize the entire transgene (B) FloxP inserted in the same direction leads to loop out genes flanked by loxP cassetts as we designed. (C). FloxP inserted in the opposite direction leads to inversion of genes flanked by loxP cassetts. (D) The principle for designing primers used for checking orientation. For head-to-tail (HT) primers, a predicted PCR product will be detected when there is at least one head-to-tail orientation. In a similar way, head-to-head (HH) primers detect head to head orientations while tail-to-tail (TT) primers detect tail-to-tail orientations. (E) The three mice lines (17, 21, and 46) contain only head-to-tail junctions. In contrast, C29-1C, which had the high numbers of transgene copies, had head-to-tail, tail-to-tail and head-to-head junctions.



**Figure 4.** Identification of transgene genomic insertion site by inverse PCR

(A) The steps for performing inverse PCR for identifying genomic location where transgene inserted. (B) Diagram illustration of genomic location of transgene insertion. The mouse genome blast results showed that the transgene insertion site on chromosome 4 is in an intergenic region between the gene encoding zinc finger protein 618, which is 308469 bp away and the gene encoding Regulator of G-protein signaling 3 isoform 3, which is 89334 bp away. The transgene insertion site is far away from the flanking genes on chromosome 4. (C) PCR verification of transgene insertion in miR-365 flox transgenic mice (TR) with forward primer recognizing genomic sequence and reverse primer recognizing transgene sequence. Four individual transgenic mice and four wild type mice genomic DNA were used for verification. Transgenic specific and non-specific bands were labelled.



**Figure 5.** MiR-365 transgene expression indicated by red fluorescence signal in different cartilaginous tissues

By breeding with col2a1-cre mice, miR-365 was conditionally up-regulated in type II collagen expressing lineage cells at different locations, including distal femur, femoral head, lower jaw, spinal column. Tissues were isolated from 7 day old miR-365 fl +/-; Col2a1-cre +/- mice. Specimens were embedded with OCT compound and cut with cryostat into 30  $\mu$ m sections. Images were taken with confocal microscope. (A) Red fluorescence indicated transgene expression in cartilage tissue at different locations. (B) Real time PCR was performed with RNAs isolated from femoral condyle and tibia plateau. miRNA-365 expression levels were compared in miR-365 fl +/-; Col2a1-cre +/- mice with cre only miR-365 fl -/-; Col2a1-cre +/- mice. (C) Femur length were measured in 2 months old transgenic mice. \* indicates  $p < 0.05$ . ( $n \geq 3$ )

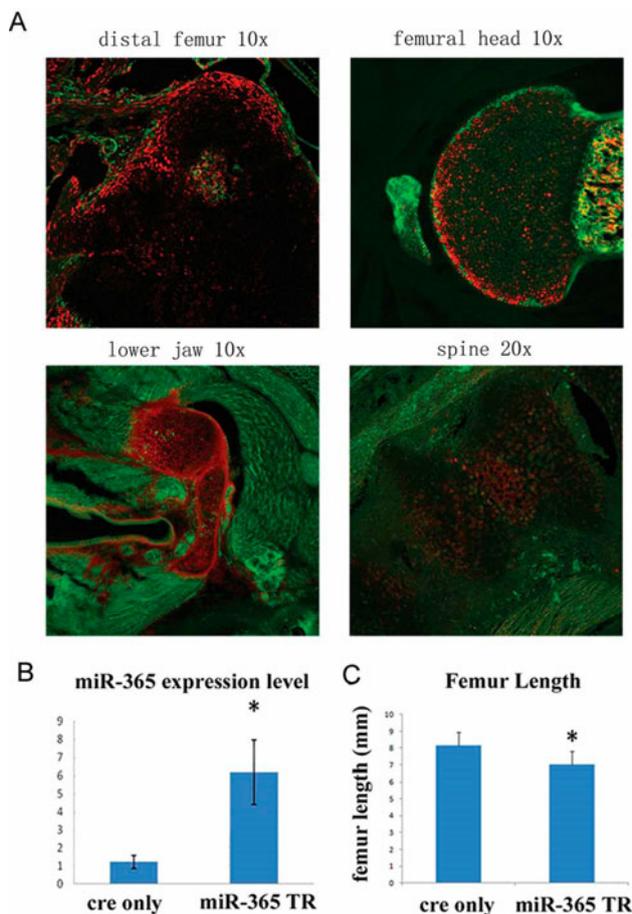


Table 1: Primers used in this study

Transgenic construct cloning	
PremiR365-1 fwd	5'- AGGACAGGTACCACCGCAGGGAAAATGAGGGACTTTTGG GGGCAGATGTG-3'
PremiR365-1 rev1	5'- GCATTATGATAGCGGAATGGAAACACATCTGCCCCCAA AGTC-3'
PremiR365-1 rev2	5'- GCCTACCTCGAGTGCAAGAGCAATAAGGATTTTTAGGGGC ATTATGATAG-3'
PCR Genotyping	
CMVE fwd	5'-ATTCGGTACC CCTGGGTCGACATTGA-3'
CMVE rev	5'-CAACGAGCTC ACCATGGTAATAGCGATG-3'
RFP fwd	5'-AGTGGGAGCGCGTGATGAACTTCGA-3'

RFP rev	5'-CTGCTCCACGATGGTGTAGTCCTCGT-3'
MiR365fl fwd	5'-AGACCCACAAGGCCCTGAAGCTGA-3'
MiR365fl rev	5'-CACATCTGCCCCCAAAGTC-3'
Transgene copy number testing with real-time PCR	
RFP-F	5'-CCGTGAAGCTGAAGGTGAC-3'
RFP-R	5'-ATCACGCGCTCCCACTTG-3'
Int-F1	5'-CTCTCTCCGTTGCAGGCAT-3'
Int-R1	5'-GTCTTACAGTGAGCTCCGTTC-3'
Transgene copy number testing	
HTF1	5'-TATTGCAGCTTATAATGGTT-3'
HTR1	5'-GGGAAATGTGCGCGGAACCC-3'

HHF1	5'-ACCGCCA ACTTGGTTGAGTACTCACCA -3'
HHR1	5'-GCGAGTGTCGTGTCAACAGCGGTAAGATCCTT -3'
HHR2	5'-CAACGCCACGTGCACGAGTGGGTTACATC -3'
TTF1	5'-ACGAATAGCCGAGCCTAGGCTTTTGCAAAAAG-3'
TTF2	5'-GTCCTCCGACATCTGAGCTATTCCAGAAGTA-3'
TTR1	5'-AGCCCAGTGCCCGTGAGGAGGCTTTTTTGGAG-3'
TTR2	5'-CGCAGTGCCCTTGCAAAAAGCTAACTTGTT-3'
TTR3	5'-CGAGTGCCGAGGCCGAGGCCGCTCGGCC -3'
Transgene insertion site testing with inverse PCR	
IPCR F1	5'- CTGCGCTCGGTCGTTCCGGCT-3'
IPCR R1	5'-TAAGATCCTTGAGAGTTTTCGCCCCGAAGA-3'

IPCR F2	5'-CAAGAGCAACTCGGTCCGCATACACTAT-3'
IPCR R2	5'-CGCGTTGCTGGCGTTTTTCCATAGGCTCCG-3'
Ch4F1	5'-GGCGGACAGGTATCCGGTAAGC-3'

Accepted Manuscript