Cell Research (2014) 24:501-504.
© 2014 IBCB, SIBS, CAS All rights reserved 1001-0602/14 $ 32.00
www.nature.com/cr

LETTER TO THE EDITOR

Rosa26-targeted swine models for stable gene over-expression and Cre-mediated lineage tracing

Cell Research (2014) **24**:501-504. doi:10.1038/cr.2014.15; published online 7 February 2014

Dear Editor,

Genetic modification of pigs has many agricultural and biomedical applications. Ectopic overexpression of foreign genes is necessary in many cases to generate transgenic pigs with favorable phenotypes [1]. However, random integration of foreign genes often leads to unpredictable expression and unstable phenotypes [2]. *Rosa26* is ubiquitously expressed in embryonic as well as adult tissues [3]. Targeting genes to the *Rosa26* locus is a desirable method to create transgenic animals consistently expressing foreign genes at a high level. Insertion of a reporter or toxin gene into the *Rosa26* locus has been widely used to trace or ablate specific cell lineages [4], and this approach plays a fundamental role in understanding cell differentiation *in vivo*. *Rosa26* was first identified and targeted in mouse embryonic stem cells (ESCs) in 1990s [3], and then in human ESCs in 2007 [5]. With the establishment of rat pluripotent stem cells, the *Rosa26* locus was also successfully identified and targeted in rats recently [6]. However, *Rosa26* has not been tackled in large animals due to unavailability of germline-competent pluripotent stem cells. By taking advantage of recently emerging technology of gene editing mediated by transcription activator–like effector nuclease (TALEN) [7], here we characterized the porcine *Rosa26* (*pRosa26*) locus and targeted a Cre-dependent reporter gene into the *pRosa26* locus. Using this approach, we also created transgenic pigs stably overexpressing a gene of interest through recombinase-mediated cassette exchange (RMCE) [5].

We identified and characterized the *pRosa26* locus based on a highly conserved promoter region of the *Rosa26* locus in mice, rats and humans [6]. We searched the Ensembl porcine database using the human *Rosa26* promoter and exon 1 sequence (1 036 bp) as a template and found a highly conserved region (sequence similarity > 90%) located on the pig chromosome 13. This region contains both the *pRosa26* locus and the neighboring genes that have also been found in mice, rats and humans (Supplementary information, Figure S1A). The sequence alignments of porcine, mouse, rat and human *Rosa26* promoter and exon 1 showed high sequence conservation (> 75%) among these species (Supplementary information, Figure S1B). In mice, rats and humans, the *Rosa26* locus encodes non-coding RNAs that are ubiquitously expressed [5, 6, 8]. To determine whether the *pRosa26* locus also encodes similar non-coding RNAs, we performed an online screening of the Ensembl porcine gene expression database for the *pRosa26* locus, but failed to identify any expressed sequence tags (ESTs) or transcripts. Therefore, we predicted the exon 1 sequence of *pRosa26* according to the multiple-sequence alignment result (Supplementary information, Figure S1B) and designed primers aligned within exon 1 to perform 3′ RACE analyses [6]. A non-coding RNA product of 629 bp transcribed from the *pRosa26* locus was identified (Supplementary information, Figure S1C and S1D). RT-PCR and SYBR green-based quantitative PCR assays demonstrated that this non-coding RNA was expressed in a wide variety of adult tissues (Figure 1A and 1B). To determine whether the *pRosa26* promoter can drive gene expression ubiquitously like the mouse promoter [8], we constructed and transiently transfected a *pRosa26* promoter-driven *tdTomato* transgene into six cell lines. At 48 h post-transfection, a high level of tdTomato expression was detected in all six cell lines, a pattern that was highly similar to the expression of the ubiquitous CMV promoter-driven *tdTomato* transgene (Figure 1C). Taken together, these data strongly suggest that the *pRosa26* locus is an ideal site for ubiquitous expression of exogenous genes.

To generate transgenic pigs with stable and ubiquitous overexpression of exogenous genes and also reliable reporters that can be used for stem cell and human disease studies, we attempted to introduce a Cre-dependent EGFP reporter into the *pRosa26* locus. As porcine ESCs for gene targeting are not available, we applied the targeting strategy to pig fetal fibroblasts (PFFs) and performed somatic cell nuclear transfer (SCNT) to generate transgenic pigs [1]. We first applied a traditional homologous recombination-mediated targeting strategy. We selected 425 colonies, but none of them was correctly targeted. Thus, the efficiency of the traditional strategy

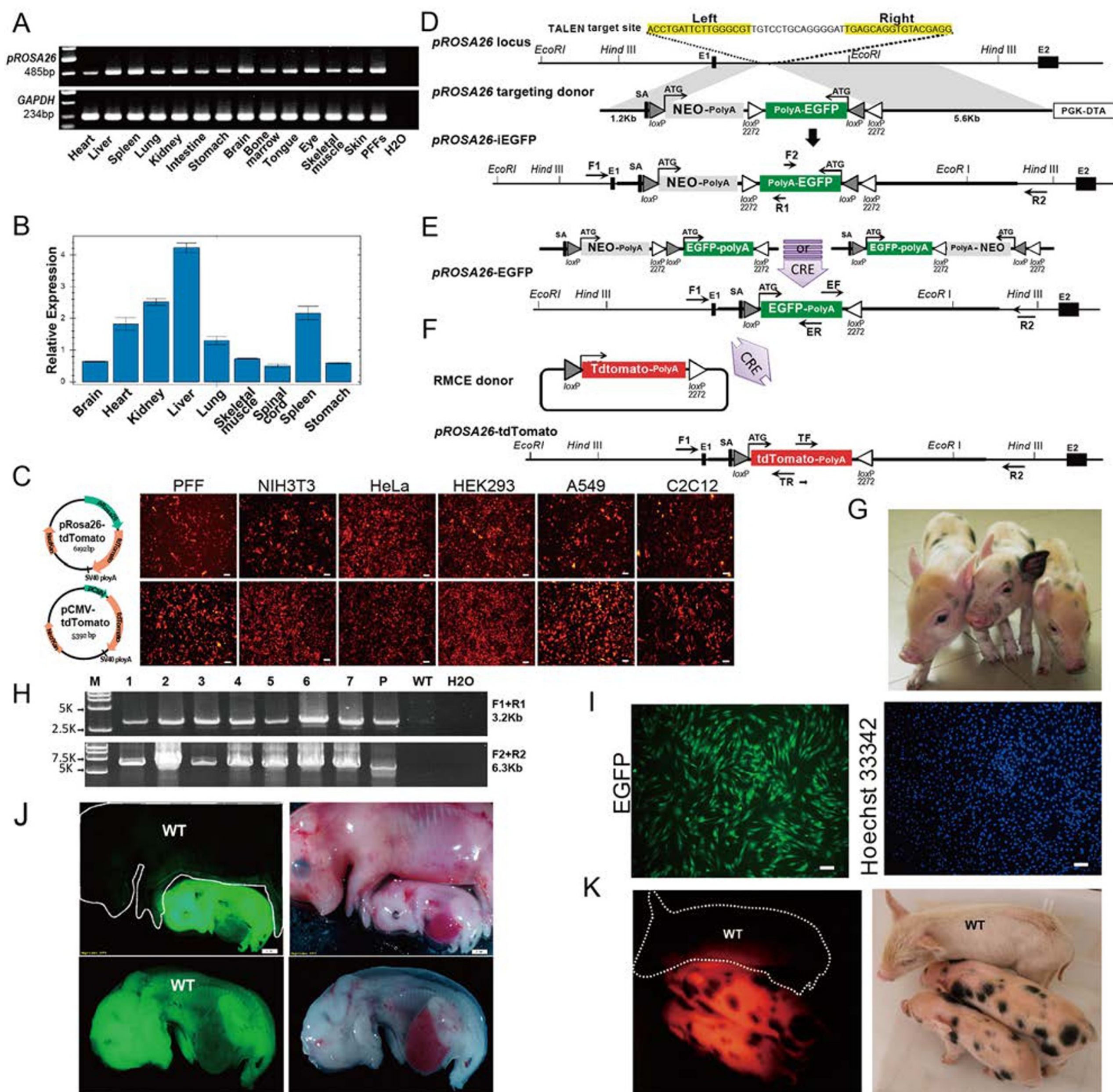

Figure 1 Characterization of *pRosa26* and highly efficient gene knock-in and replacement at the *pRosa26* locus. **(A**, **B)** *pRosa26* was expressed in a variety of organ tissues as determined by RT-PCR **(A)** and quantitative RT-PCR **(B)**. For RT-PCR, the designed primers annealed in exon 1 and exon 2 and amplified a correctly spliced product of 485 bp. Porcine *GAPDH* was used as a control (234 bp). For qPCR, primers were specific for exon 2. PCR product of the porcine *ACTB* gene was used as the reference control. Data were presented as the average expression levels from three individual RT/qPCR reactions. **(C)** *pRosa26* promoter-driven tdTomato expression in different cell lines (*pRosa26* promoter region was shown in Supplementary information, Figure S1B and S1C). Transient transfection of *pRosa26*-tdTomato and pCMV-tdTomato vectors was performed in the indicated cell lines. **(D)** A diagram for TALEN-mediated knock-in of *Neo-polyA-iEGFP* into the *pRosa26* locus. Grey triangles, wild-type (WT) *lox*P; white triangles, *lox*P2272 site; SA, splice acceptor. **(E)** Cre-mediated recombination activates EGFP expression by two mechanisms: Cre induces inversion of the *iEGFP* flanked by two *lox*P2272 sites (upper left) followed by excision of *Neo* flanked by two *lox*P sites; or Cre induces inversion of both *Neo* and *iEGFP* flanked by two *lox*P sites (upper right) followed by excision of *Neo* between two *lox*P2272 sites. **(F)** RMCE replaces *EGFP* with *tdTomato*. **(G)** Morphologically normal piglets were born from SCNT with the *pRosa26-iEGFP* PFFs. **(H)** PCR analysis confirmed the correct homologous recombination at the *pRosa26* locus in the 7 piglets generated by SCNT. The *pRosa26-iEGFP* donor cells were used as positive control (P) and WT pig genomic DNA and water were used as negative controls. Primer pairs were shown in **D**. **(I)** EGFP activation by Cre in fibroblasts isolated from the ear tissues of cloned piglets shown in **G**. Cells were infected with Cre-lentivirus and the EGFP expression was observed after 48 h. **(J)** Embryos with constitutively activated EGFP expression were generated from SCNT of the *pRosa26*-EGFP PFFs obtained with a Cre plasmid transient transfection. Shown were an E35 *pRosa26*-EGFP embryo and its section. WT is an E45 WT embryo. **(K)** SCNT-generated *pRosa26*-tdTomato piglets through an RMCE strategy as shown in **F**.

appeared too low for us to obtain targeted clones.

We then applied the TALEN technology to improve the targeting efficiency. Six TALENs (Supplementary information, Figure S2A) composed of 14.5-16.5 repeats were designed based on published guidelines [7]. The activity of the TALENs was tested using a single-strand annealing (SSA) assay [9] (Supplementary information, Figure S2B and S2C) and a T7 endonuclease I (T7EI) assay [7] (Supplementary information, Figure S2D and S2E). The TALEN pairs with the highest activity (Figure 1D and Supplementary information, Figure S2F) were used to target the *pRosa26* locus. The targeting vector (Figure 1D) contains an 1.2 Kb 5′ arm and a 5.6 Kb 3′ arm for homologous recombination. These homologous arms span 6.8 Kb of the *pRosa26* locus containing the putative first exon and the following intron of *pRosa26*. An expression cassette comprising a viral splice acceptor (SA), a promoterless neomycin-resistance (*Neo*) gene, and an inverted *EGFP* (*iEGFP*) gene was inserted between the homologous arms. The *lox*P and mutant *lox*P2272 sites were arranged to flank the *Neo* and *iEGFP* genes as indicated in Figure 1D, which could result in removal of the *Neo* gene and inversion of the *iEGFP* after Cre-mediated recombination. This Cre-mediated rearrangement would place *EGFP* directly under the control of the endogenous *pRosa26* promoter (Figure 1E). Another advantage of putting heterotypic *lox*P sites in the knock-in vector is that virtually any gene of interest could be inserted into the *pRosa26* locus by RMCE.

For gene targeting, PFFs were electroporated with either the linear or circular targeting construct accompanied with TALEN pairs. After selection with G418 (1 mg/ml from day 10 to day 14), 192 cell clones were expanded and screened by PCR analysis. A total of 46 out of 96 clones (*pRosa26*-iEGFP) derived from linear construct electroporation and 14 out of 96 clones obtained from circular construct electroporation were correctly targeted based on 5′- and 3′-arm PCR analysis and EGFP expression induced by Cre-lentivirus infection (Figure 1D and Supplementary information, Figure S3A-S3E). The TALEN generates site-specific DNA double-strand breaks (DSBs), which can be repaired by nonhomologous end joining (NHEJ) or homology-directed repair [7]. Once DSBs are erroneously repaired by NHEJ, unpredictable gene mutations will be generated. Therefore, we next examined the modification patterns in both *pRosa26* alleles among these 60 correctly targeted cell clones using PCR-based DNA sequencing. We found that 39 clones had knock-in mutation in one allele and NHEJ-mediated mutation in the other, but no homozygous knock-in clones were identified. We chose the clones with only one *Neo*-*iEGFP* knock-in *pRosa26* allele but no NHEJ-mediated mutation for further validation. These cells exhibited a normal karyotype (Supplementary information, Figure S3F) and expressed a high level of EGFP when infected with the Cre-lentivirus (Supplementary information, Figure S3C and S3D). Two correctly targeted PFF cell lines obtained from the linear and circular constructs, respectively, were chosen as donor cells for SCNT. A total of 1 088 cloned embryos were generated and transferred into four surrogate mothers. Two surrogates were confirmed pregnant by ultrasound examination one month after the embryo transfer.

To verify whether the expression of EGFP in the cloned *pRosa26*-iEGFP fetuses could be induced by the Cre-recombinase, a pregnant surrogate was sacrificed at day 35 after the embryo transfer. Six fetuses were collected to isolate PFFs. The PFFs were transiently transfected with a Cre expression plasmid (Figure 1E) and the expression of EGFP in those cells were observed 48 h later. EGFP-expressing (*pRosa26*-EGFP) cells were sorted by flow cytometry. PCR analysis confirmed the correct removal of the *Neo* gene and the inversion of *iEGFP* in the *pRosa26*-EGFP cells (Supplementary information, Figure S4A). Karyotype analysis showed that these cells contained normal chromosomes (Supplementary information, Figure S4B). The other surrogate mother was allowed to develop to term and delivered seven *pRosa26*-iEGFP piglets (three live piglets (Figure 1G) and four stillbirths). PCR analysis showed that all the seven piglets were derived from the SCNT donor cells (Figure 1H). We also examined whether the cloned piglets could express EGFP after Cre-mediated recombination. Fibroblasts isolated from the ear tissues of the live piglets were infected with Cre-lentivirus. As shown in Figure 1I, a high level of EGFP expression was observed in these cells (Figure 1I).

To further examine the expression pattern of EGFP *in vivo* after Cre-mediated recombination, *pRosa26*-EGFP PFF donor cells were used for the second SCNT. The pregnancy of a surrogate was terminated 35 days after the embryo transfer and six fetuses were obtained. EGFP was indeed ubiquitously expressed in all organs of the six fetuses (Figure 1J). In addition, flow cytometry analysis revealed that all fibroblasts isolated from the fetuses expressed EGFP (Supplementary information, Figure S4C and S4D). All the results above indicate that the *pRosa26* locus is an excellent site for ubiquitous expression of exogenous genes. In addition, the Cre-inducible *pRosa26*-iEGFP reporter line could serve as an ideal tool to trace cell lineages in pigs.

The insertion of heterotypic *lox*P sites along with *EGFP* into *pRosa26* introduced a homing site (Figure 1E), which could allow us to replace the EGFP cassette at the *pRosa26* locus with any gene of interest by RMCE without drug selection. To test the feasibility of this ap-

proach, we engineered an exchange vector (Figure 1F) containing a promoterless *tdTomato* flanked by 5′ *lox*P and 3′ *lox*P2272 sites. This exchange vector, together with a Cre expression plasmid (Figure 1F), was electroporated into *pRosa26*-EGFP fetal fibroblasts. Forty-eight hours after electroporation, tdTomato-expressing cells started to be observed (Supplementary information, Figure S5A). And tdTomato$^+$/EGFP$^-$ (*pRosa26*-tdTomato) cells were sorted 5 days later. PCR analysis further confirmed the successful replacement of *EGFP* with *tdTomato* (Supplementary information, Figure S5B). Karyotype analysis showed that the sorted cells contained normal chromosomes (Supplementary information, Figure S5C). Again, *pRosa26*-tdTomato cells were used as donor cells for SCNT, and four *pRosa26*-tdTomato piglets (2 live piglets and 2 stillbirths) were obtained. tdTomato fluorescence was directly observed in piglets using goggles [10] (Figure 1K). Flow cytometry analysis showed that tdTomato was expressed in all *pRosa26*-tdTomato piglet ear fibroblasts (Supplementary information, Figure S5D and S5E) and a large majority (98.5%) of blood cells with nuclei (Supplementary information, Figure S5F). To examine the expression pattern of tdTomato *in vivo*, one piglet was sacrificed and tdTomato expression was observed in all the organs examined under a fluorescence stereomicroscope (Supplementary information, Figure S5G and S5H). Quantitative PCR assays showed that the expression pattern of tdTomato was similar to that of the *pRosa26* non-coding RNA in various adult tissues (Supplementary information, Figure S5I). Taken together, these results indicate that transgenic pigs stably overexpressing a gene of interest could be generated through RMCE in the *pRosa26* locus without drug selection. The stable and ubiquitous expression of *EGFP* or *tdTomato* at the targeted *pRosa26* locus is in stark contrast to the expression of tdTomato in a pCMV-tdTomato transgenic pig that we generated through random integration. In that case, only 18% of pig ear-derived fibroblasts expressed tdTomato and the expression levels also varied significantly among individual cells (Supplementary information, Figure S6).

In summary, we have for the first time identified and characterized the porcine *Rosa26* locus. Similar to the findings in mice, rats and humans, *pRosa26* was also expressed in all tissues examined and the *pRosa26* promoter was capable of driving gene expression in all cell lines tested. We genetically manipulated the locus and created a Cre-inducible EGFP reporter pig line, which could be used as a reliable porcine reporter model for lineage tracing studies. We further confirmed that foreign genes inserted in the *pRosa26* locus could be expressed ubiquitously. More importantly, the RMCE technology could be used to produce transgenic pigs stably overexpressing any gene of interest at the *pRosa26* locus. The *pRosa26*-targeted pigs, together with the RMCE strategy reported in this study, will serve as an excellent platform for generating genetically modified pigs for biomedical and agricultural applications.

Acknowledgments

This work was supported by grants from the National Basic Research Program of China (973 programs; 2011CB944203), the National Natural Science Foundation of China (31272394), the Program for Changjiang Scholars and Innovative Research Team in University (PCSIRT; IRT1248), the Special Major Program for Transgenic Organisms of China (2014ZX08006-002B), and the Science and Technology Program of Guangzhou (12S134060176 and 2010U1-E00811-5) to LL and by grants from the National Institutes of Health (HL109054) and the Samuel and Jean Frankel Cardiovascular Center, University of Michigan (Inaugural Fund) to WZ.

Xiaoping Li[1, *], Yi Yang[1, *], Lei Bu[4], Xiaogang Guo[2], Chengcheng Tang[1], Jun Song[2, 3], Nana Fan[2], Bentian Zhao[2], Zhen Ouyang[2], Zhaoming Liu[2], Yu Zhao[2], Xiaoling Yi[2], Longquan Quan[1], Songcai Liu[1], Zhenguo Yang[1], Hongsheng Ouyang[1], Y Eugene Chen[3], Zhong Wang[3], Liangxue Lai[1, 2]

[1]Jilin Provincial Key Laboratory of Animal Embryo Engineering, Institute of Zoonosis, College of Veterinary Medicine, Jilin University, Changchun 130062, China; [2]Key Laboratory of Regenerative Biology, Chinese Academy of Sciences, Guangdong Provincial Key Laboratory of Stem Cells and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Guangzhou, Guangdong 510530 China; [3]Department of Cardiac Surgery, Cardiovascular Research Center, University of Michigan, 2800 Plymouth Road, Ann Arbor, MI 48109, USA; [4]Leon H Charney Division of Cardiology, New York University School of Medicine, 522 First Avenue, New York, NY 10016, USA

*These two authors contributed equally to this work.
Correspondence: Liangxue Lai[a], Zhong Wang[b]
[a]E-mail: lai_liangxue@gibh.ac.cn
[b]E-mail: zhongw@med.umich.edu

References

1 Lai L, Kang JX, Li R, *et al*. *Nat Biotechnol* 2006; **24**:435-436.
2 Yang D, Wang CE, Zhao B, *et al*. *Hum Mol Genet* 2010; **19**:3983-3994.
3 Friedrich G, Soriano P. *Genes Dev* 1991; **5**:1513-1523.
4 Dulauroy S, Di Carlo SE, Langa F, *et al*. *Nat Med* 2012; **18**:1262-1270.
5 Irion S, Luche H, Gadue P, *et al*. *Nat Biotechnol* 2007; **25**:1477-1482.
6 Kobayashi T, Kato-Itoh M, Yamaguchi T, *et al*. *Stem Cells Dev* 2012; **21**:2981-2986.
7 Reyon D, Tsai SQ, Khayter C, *et al*. *Nat Biotechnol* 2012; **30**:460-465.
8 Zambrowicz BP, Imamoto A, Fiering S, *et al*. *Proc Natl Acad Sci USA* 1997; **94**:3789-3794.
9 Wilson KA, Chateau ML, Porteus MH. *Mol Ther Nucleic Acids* 2013; **2**:e87.
10 Deng W, Yang DS, Zhao BT, *et al*. *PLoS One* 2011; **6**:e19986.

(**Supplementary information** is linked to the online version of the paper on the *Cell Research* website.)