Construction and identification of Rab7-ROSA26 knock-in mice by CRISPR/Cas9 technology

Qing Xu

Affiliated hospital of jiangnan university <https://orcid.org/0000-0002-9964-3983>

Linfang Jin

Affiliated hospital of jiangnan university

Xiaowei Qi (✉ 295518165@qq.com)

<https://orcid.org/0000-0001-6650-3629>

Methodology article

Keywords: CRISPR/Cas9, Genome editing, Rab7, Renal fibrosis

Posted Date: August 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-58060/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: CRISPR/Cas9 system, a Genome editing technique, can achieve the purpose of gene knockout and gene insertion. Because we want to study the relationship between Rab7 and renal fibrosis in vivo, we should construct Rab7-knock-in mice.

Results: We constructed Rab7-knock-in mice by CRISPR/Cas9 technology. Using the identification of PCR, southern blot and sequencing, the detected results were in line with our expectation.

Conclusions: We constructed Rab7-knock-in mice successful, which can be provided for studying the relationship between Rab7 and renal fibrosis.

Background

Genome editing is a technique to modify the genome at a specified point and obtain the expected genome sequence of the organism for producing genetic changes[1]. It mainly involves DNA sequence knock-out, insertion, fixed point mutation and combinatorial editing. Genome editing technology has increasingly become a hotspot in biological research. Currently, genome editing technology mainly includes macroribozyme technology, zinc-finger nuclease technology, transcriptional activator effector nuclease technology and CRISPR related nuclease technology.

The CRISPR/Cas system is a prokaryotic immune system designed to resist invasion of foreign genetic material, such as bacteriophage viruses and foreign plasmids. It also provides an acquired immunity to bacteria which is similar to the secondary immunity of mammals. When bacteria are invaded by viruses or foreign plasmids, it will produce a corresponding "memory" that can resist their reinvasion.

CRISPR related nuclease technology only needs a small guide RNA to identify specific DNA, and a new site only needs a new gRNA, which greatly simplifying the process of genome editing, expanding the selection of target sites, and having greater potential. The CRISPR system can also be used for targeted gene mutations and gene modification, aimed to the deletion of large fragments of DNA and complex gene editing.

Because we wanted to study the relationship between Rab7 and renal fibrosis in vivo, we constructed Rab7-knock-in mice by CRISPR/Cas9 technology, using PCR, southern blot and sequencing to identify them.

Results

Selection of the targeting vector

Targeting Strategy of mouse Rab7-ROSA26 knock-in project was shown as (Fig. 1A). Mouse genomic fragments containing homology arms were amplified from BAC clone by using high fidelity Taq, and were sequentially assembled into a targeting vector together with recombination sites and selection markers shown as (Fig.1B). The targeting vector was digested by restriction enzymes for confirmation purposes (Fig.1C). And sequence of the final targeting vector were shown as (Fig. 2, 3).

The identification of PCR and southern blot

The gRNA to mouse ROSA26 gene, the donor vector containing mouse Rab7 cDNA, and Cas9 were coinjected into fertilized mouse eggs to generate targeted knockin offspring. F0 founder animals were identified by PCR followed by sequence analysis, which were bred to wildtype mice to test germline transmission and F1 animal generation. The genotyping strategy of PCR screening was shown as (Fig.4A). F1 animals 21, 22, 23, 26 and 28 were identified positive by PCR screening for region 1 and region 2 (Fig.4B). The correct gene targeting in 5 F1 animals (21, 22, 23, 26 and 28) were confirmed by Southern blot analysis of the tail DNA samples. The strategy of Southern blot analysis is shown as (Fig.4C). Expected Fragment Sizes for Southern Blot: (5'Probe-Sspl: 4.09 kb-WT, 7.03 kb-MT/ 3'Probe-Mfel: 12.61 kb-WT, 10.39 kb-MT). All of the five mice (21, 22, 23, 26 and 28) were confirmed correct by Southern blot analysis (Fig.4D).

Sequencing confirmation of Rab7-knock-in mice

The diagram of sequencing confirmation was shown as (Fig.5A). And as one example of positive mice, the mouse Rab7 cDNA was knocked-in successfully (Fig. 5B).

Dicussion

CRISPR cluster is a family of special DNA repeats widely existing in the genomes of bacteria and archaea. It is a cluster of short palindromic repeats at regular intervals, distributed among 40% of sequenced bacteria and 90% of sequenced archaea. The CRISPR sequence consists of a number of short and conservative repeats and spacers. Repeat sequence region contains palindrome sequence, which can form hairpin structure. The spacers are special, they are the foreign DNA sequences captured by bacteria. This is the equivalent of a "blacklist" of the bacterial immune system. CRISPR/Cas system strikes with precision when the foreign genetic material invades again. The leader in the upstream is considered to be the promoter of the CRISPR sequence. In addition, there is a polymorphic family gene that can interact with the CRISPR sequence region. Therefore, the gene is named CRISPR associated gene (Cas). The Cas gene coevolve with the CRISPR sequence to form a highly conserved CRISPR/Cas system in bacteria.

Actually, CRISPR/Cas9 technology takes three steps to destroy the foreign DNA.

Firstly, the CRISPR/Cas system implements a "blacklist registration" function in this step. The CRISPR/Cas system identifies the intruder's "name" (PAM) and finds its "id card" (original interval sequence), and then records the intruder's identity information as a "file" (interval sequence) to the "blacklist" (CRISPR sequence). When a phage virus invades a host bacterium, the virus's double-stranded DNA is injected into the cell. The CRISPR/Cas system intercepts a sequence from this foreign DNA as an "identity card" of the foreign DNA and integrates it into the genomic CRISPR sequence as a new interval sequence. Therefore, the identity card corresponding to the interval sequence is called protospacer. However, the selection of "id card" is not random. Several bases extending to both ends of the proto-spacer sequence are conservative and are called proto-spacer adjacent motif (PAM). When the virus invades, the proteins encoded by Cas will scan the foreign DNA and identify the PAM region, and then take the DNA sequence adjacent to PAM as the candidate original interval sequence. Subsequently, the Cas protein cuts the original spacer sequence from the foreign DNA and inserts the original spacer sequence into the downstream of the adjacent leading region of the CRISPR sequence with the assistance of other enzymes. In this way, a new interval sequence is added to the CRISPR sequence of the genome.

Secondly, some special structures are needed to destroy the foreign DNA. When foreign DNA invades, the CRISPR sequence transcribes two kinds of RNA under the regulation of the promoter in the leading region, named pre-crRNA and tracrRNA. TracrRNA is an RNA with hairpin structure transcribed from repetitive sequence region, while pre-crRNA is a large RNA transcribed from the whole CRISPR sequence. Then pre-crRNA, tracrRNA and proteins encoded by Cas9 are assembled into a complex. It will select the corresponding "identity card", and cut the sequence with the help of RNase III. Eventually a complex consisting of crRNA, Cas9, and tracrRNA is formed.

Finally, it's the process of targeted interference. This complex will scan the entire foreign DNA sequence and identify the original spacer sequence that is complementary to crRNA. Then, the complex locates the region of the PAM and the DNA double strand is unwound, forming the R-loop. The crRNA will hybridize to the complementary strand, while the other strand remains free. Cas9 protein can cut the target DNA by two nuclease domains within the molecule, each of which is responsible for cutting a DNA strand in the R-loop of the target DNA. In the end, the Cas9 protein breaks the double strand and the expression of the foreign DNA is silenced.

CRISPR/Cas is a powerful tool for gene editing, allowing precise and targeted editing of genes. With the participation of guide RNA and Cas9, the cell genome DNA to be edited will be treated as a virus or foreign DNA. But there are limits to the use of CRISPR/Cas9. A relatively conservative PAM sequence is necessary to exist near the region to be edited. And guide RNA is complementary to the sequence bases upstream of PAM. The most basic technique is gene knock-out. If two guide RNAs are designed in the upstream and downstream of the gene, and they are transferred into the cell together with the plasmid containing Cas9, the guide RNA can target the target sequence near PAM through base complementary pairing, and Cas9 will break the DNA double strand in the upstream and downstream of the gene. The organism itself has a response mechanism of DNA damage repair, which connects the sequences at the upper and lower ends of the break, thus achieving the knockout of the target gene in the cell. If a repair

template plasmid (donor DNA) is introduced to the cell on this basis, then the cell will introduce fragment insertion or fixed point mutation in the repair process according to the template provided. This allows for gene replacement or mutation. Gene editing of fertilized egg cells and introducing them into the surrogate mother can construct the gene editing animal model.

However, the commonly used CRISPR/Cas9 system still has several defects. It does not have strict matching requirements for guide RNA, and Cas9 may not follow the instructions of guide RNA for random cutting, resulting in missed editing and even large segment loss at the remote editing site. But A.V. Anzalone et al.[2] improved these problems, on the one hand, they modified Cas9 protein so that it could only cut off the double-stranded DNA in a chain. On the other hand, a template RNA of DNA was added to connected with guide RNA, which became the editing extension guide RNA(pegRNA). A reverse transcriptase was also added to the Cas accordingly, and the DNA from the reverse transcription of the template RNA was used to repair the target site. Thus, the problems of editing miss and insertion deletion were greatly improved.

With the deepening and comprehensive research on CRISPR/Cas9 technology, it will be more accurate and easy to construct transgenic animals to study human diseases, for making greater contributions to human health in the future.

Conclusion

Our work constructed the Rab7-knock-in mice successfully by CRISPR/Cas9 technology. And It provided the model to us for studying the relationship between Rab7 and renal fibrosis.

Methods

Animals

WT C57BL/6 mice were purchased from Maijiesi (Nanjing, China), gRNA was specifically introduced into the ROSA26 locus of C57BL/6 mice. Together with a plasmid containing mouse Rab7 cDNA and cas9, the gRNA was injected into fertilized mouse eggs by prokaryotic microinjection to establish Rab7-knock-in mice. After experimentation, the WT mice were euthanized after sedated by 4% chloral hydrate (350 mg/kg).

DNA extraction and PCR

We used TaKaRa MiniBEST Universal Genomic DNA Extraction kit to gain high purity of genomic DNA. **a.** Add 180 μ L of Buffer GL, 20 μ L of Proteinase K and 10 μ L of RNase A per tail piece (2-5 mm) in a microcentrifuge tube. **b.** Incubate the tube at 56°C overnight. **c.** Spin in microcentrifuge at 12,000 rpm for 2 minutes to remove impurities. **d.** Add 200 μ L Buffer GB and 200 μ L absolute ethyl alcohol with sufficient mixing. **e.** Place the spin Column in a collection tube. Apply the sample to the spin and

centrifuge at 12,000 rpm for 2 min, discard flow-through. **f.** Add 500 µL Buffer WA to the spin column and centrifuge at 12,000 rpm for 1 min, discard flow-through. **g.** Add 700 µL Buffer WB to the spin column and centrifuge at 12,000 rpm for 1 min, discard flow-through. **h.** Repeat step 4. **i.** Place the spin Column in a collection tube and centrifuge at 12,000 rpm for 2 min. **j.** Place the spin Column in a new 1.5ml tube. Add 50~200 µL sterilized water or elution buffer to the center of the column membrane and let the column stand 5min. **k.** To elute DNA, centrifuge the column at 12,000 rpm for 2 min. To increase the yield of DNA, add the flow-through and/or 50~200 µL sterilized water or elution buffer to the center of the spin column membrane and let the column stand 5 min. Centrifuge at 12,000 rpm for 2 min. **l.** Quantify to genomic DNA. Eluted genomic DNA can be quantified by electrophoresis.

PCR Primers for Region 1: 5'arm forward primer (F1): 5'-AATTCCTGTGTCCTCCAAACTCAG-3'/ 3'KI reverse primer (R1): 5'-AGATGTACTGCCAAGTAGGAAAGTC-3'. PCR Primers for Region 2: 5'KI forward primer (F2): 5'-GGTTGGCTATAAAGAGGTCATCAG-3'/ 3'arm reverse primer (R2): 5'-TCCAAACTGCATAGCAACATTTAACACAG-3'.

Southern Blot

a. DNA molecule. **b.** DNA was prepared by agarose gel electrophoresis. **c.** Put the gel into HCl for purine removal, when the bromophenol blue in the gel turns yellow, take it out. **d.** Wash it twice with ddH₂O, and then put it into NaOH for 45 min. **e.** Place the rubber bottom upward on the filter paper. **f.** Cut the nylon film of the same size as the glue and lay it on the glue. **g.** Cover the nylon membrane with plastic cloth, cut a piece of filter paper of the same size as the nylon membrane, and lay it on the nylon membrane. **h.** Put it overnight. **i.** After the transfer, mark the cut angle of nylon film, put it on a piece of filter paper, UV cross-linking 30s. **j.** Pre hybridization. **k.** Hybridization. **l.** Recover the hybrid liquid and freeze it, 14. Add 50 ml 0.5 × SSC / 0.1% SDS preheated at 66 °C into the hybrid tube, wash at 66 °C for 30 min. **m.** Add washing buffer and soak for 5 min. **n.** Add blocking buffer. **o.** Add antibody solution, and incubated at 25 °C for 1 h. **p.** Take the nylon film out of the hybrid tube and put it into the large culture with detection buffer, soak for 5 minutes. **q.** Add CSPD solution evenly, 37 °C incubation for 10 min. **r.** Press the film for 2 h. **s.** Film developing.

Primers for 5'Probe: 5'Probe forward primer: 5'-AAACCAGGGCTTTTGCCCACTAC-3', 5'Probe reverse primer: 5'-TTGTGACAACTCTGGTCGCTCGT-3'. Primers for 3'Probe: 3'Probe forward primer: 5'-TTCTGGGCAGGCTTAAAGGCTAAC-3', 3'Probe reverse primer: 5'-AGGAGCGGGAGAAATGGATATGAAG-3'.

Sequencing

Sequencing Primer for Region 1 PCR product: 5'Sequence primer (F3): 5'-CACTTGCTCTCCCAAAGTCGCTC-3'. Sequencing Primer for Region 2 PCR product: 3'Sequence primer (R3): 5'-ATACTCCGAGGCGGATCACAA-3'.

Abbreviations

gRNA: guide RNA; WT: Wild Type; PAM: Proto-spacer Adjacent Motif

Declarations

Ethics approval and consent to participate

This study was approved by the Animal Research Ethics Committee of the School of Medicine, Southeast University(N0.20171202006).

Consent for publication

Not applicable.

Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81370868). The funder was the corresponding author of this article.

Authors' contributions

XQ conceived, designed and performed the experiments. XQ, JLF, QXW analyzed the data. XQ wrote the paper. All authors read and approved the final manuscript.

Acknowledgments

Not applicable.

References

1. Williams A, Henao-Mejia J, Flavell RA. Editing the mouse genome using the CRISPR-Cas9 system. Cold Spring Harbor Protoc. 2016;pdb.top087536.
2. Anzalone AV, Randolph PB, Davis JR, Sousa AA, Koblan LW, Levy JM, et al. Search-and-replace genome editing without double-strand breaks or donor DNA. Nature. 2019;576:149-157.

Figures

Figure 1

Selection of the targeting vector. A. The targeting strategy. B. The diagram of linearized targeting vector. C. PCR results of enzyme digestion of final vector. (1. AhdI : 6.0/4.8/3.3) (2. DrdI : 7.7/3.2/1.9/1.4) (3. NcoI : 5.0/4.8/2.5/1.3/0.6) (Kbp).

1 CTAATA TGTAAG CGTTAA TATTTT GTTAAA ATTCCG GTTAAA TTTTGT TTAATA CAGCTC
 61 ATTTT TAACCA ATAGGC CGAAAT CGGCAA AATCCC TTATAA ATCAAA AGAATA GACCGA
 121 GATAGG GTTGAG TGTGTG TCCAGT TTGGAA CAAGAG TCCACT ATTAAG GAACGT GGACTC
 181 CAACGT CAAGAG GCGAAA AACCGT CTATCA GGGCGA TGGCCC ACTACG TGAACC ATCACC
 241 CTAATC AAGTTT TTTGGG GTCGAG GTGCGG TAAAGC ACTAAA TCGGAA CCTAA AGGAGG
 301 CCCCCG ATTTAG AGCTTG ACGGGG AAAGCG GGGCGA CGTGCG GAGAAA GGAAGG GAAGAA
 361 AGCGAA AGGAGC GGGCGC TAGGGC GCTGGC AAGTGT AGCGGT CACGCT GCGCGT AAACCA
 421 CACACC CCGCGC GCTTAA TGGCGC GCTACA GGGCGC GTCCCA TCGCCG ATTCAG GCTGCG
 481 CAACTG TTGGGA AGGGCG ATCGGT GGGGGC CTCTTC GCTATT ACGCCA GCTGGC GAAAGC
 541 GGGATG TGCTGC AAGGCG ATTAAG TTGGGT AACGCC AGGGTT TTCCA GTACGC ACGTTG
 601 TAAACA GACGCG CAGTGA ATTGTA ATACGA CTCACT ATAGGG GAAATT GGGTAC GCGGGC
 661 GCATTC TGGTAC CCCTCA GCCCTA AGAATG AGAAGG GCAAAG GCGCCC GATAGA ATAAAT
 721 TAGCAT AGAAGG GGTCTT CCCAGG AGTTAA AACTTT CCTTCT GAGCGA TTAECT ACTAAA
 781 ACCAGG GCTTTT GCCCAC TACCAT TTACTT AGGATT TTGGCT TGCACG GATTCA TAGGGG
 841 CATATC CTTCCC CTTCTT CTITAG AGTCGT TCTTAA AAGATC GCTCTC CACGCC CTAGGC
 901 AGGGAA AACGAC AAAATC TGGCTC AATCCC AGGCTA GAACCC TACAAA TTCAAC AGGGAT
 961 ATCGCA AGGATA CTGGGG CACACG CCACAG GAGTCT CAAGAA TGTGAG GTGGGG GTGGCG
 1021 AAGGTA ATGTCT TTGGTG TGGGAA AAGCAG CAGCCA TCTGAG ATAGGA ACTGGA AAACCA
 1081 GAGGAG AGGCGT TCAGGA AGATTA TGGAGG GGAGGA CTGGGC CCCCCA GAGCGA CCAGAG
 1141 TTGTCA CAAGGC CGCAAG AACAGG GGGAGT GGGGGG CTCAGG GACAGA AAAAAA AGTATG
 1201 TGTATT TTGAGA GCAGGG TTGGGA GGGCTC TCTGTA AAAGGG TATAAA CTGTGA GTAGCG
 1261 AATACC CAGGCA AAAAGG GGAGAC CAGAGT AGGGGG AGGGGA AGAGTC TGAACC CAGGGG
 1321 AGACAT TAAAAA GGTAGT GGGGTC GACTAG ATGAAG GAGAGC CTTTCT CTCTGG GCAAGA
 1381 GCGGTG CAATGG TGTGTA AAGGTA GCTGAG AAGACG AAAAGG GCAAGC ATCTTC CTGCTA
 1441 CCAGGC TGGGGA GGCCCA GGCCCA CGACCC CGAGGA GAGGGA ACGCAG GGAGAG TGAAGT
 1501 GACCCT TCTTTC CCCCAG GGGCGG GTCTGT TGGTTC GGTGTC TCTTTT CTGTG GACCCT
 1561 TACCTT GACCCA GCGGCT GCGGGG GCTTGG GCGCGC GTCTGC GCGCAC GGCCTT CCGGGG
 1621 AGGCAG CGAGAG TCGAGT TAGGCC CAACGC GCGGCC ACGGCG TTCTCT GGGCGG GAATGG
 1681 CCGGTA CCGGTG AGGTGG GGGTGG GGGGCA GAAAGG GCGGAG CGAGCC CGAGGC GGGGAG
 1741 GGGAG GGGCAG GGGCGG AGGGGG CCGGCA CTACTG TGTGGG CCGACT GCGGGG ACTAGG
 1801 GCTGCG TGAGTC TCTGAG CGCAGG CCGGCG GCGGCC GCGCCT CCGCGG GCGGCG GCAAGC
 1861 GCGGCA GCGGCG GCAGCT CACTCA GCGCGC TGCGCG AGCGGA AACGCC ACTGAG CGCAGC
 1921 GGGATT CCGAGT GCGGCG GCCAGG GGCACG CCGGAC ACGGCC CTTCCC GCGGCG CCATTG
 1981 CCGTCT CCGCCC ACCGCG CCACAC TTATTG GCGGGT GCGGCG CCAATC AGCGGA GGCCTG
 2041 CCGGGC CGCTCA AAGAAAG AGGCTG TGTCTT GGGGCT CCGGCT CTTGAG AGAGGC TCGGCT
 2101 AGGTAG GGGATC GGGACT CTGGCG GGAGGG CCGGCT GGTGCG TTGCGG GGGATG GGGGCG
 2161 CCGGCG AGGCCC TCCGAG CGTGGT GGAGCC GTTCTG TGAGAC AGCGCG GTACGA GTGCTG
 2221 ACGCTG GAAAGG GCAAGC GGGTGG TGGGCA GGAATG CGGTCC GCGCTG CAGCAA CCGGAG
 2281 GGGGAG GGAAGG GGGAGC GGAAGG GTCTCC ACCGGA CCGGCG CATGGC TCGGGG GGGGGG
 2341 GGGCAG CGAAGG AGCGCT TCGGCG CGAGCT CTGCTC GCTGAT TGGCTT CTTTTC CTTCCC
 2401 CCGTGT GTGAAA ACACAA ATGGCG GTTTTT GGTGGG CGTAAG GCGCCT GTCAAT TAACGG
 2461 CAGCGG GAGTGC GCAGCC GCGGCG AGCCTC GCTCTG CCAACT GGGTGG GCGGCG AGGTAG
 2521 GTGGGG TGAGGC GAGCTG GACGTG CCGGCG CCGTGC GCTCTT GCGGCG GCGGCG GAGGGG
 2581 AGGGAG GGTGAG CGAAAG TAGCTC GCGGCG GAGGCG CCGCCC ACCCTC CCTTTC CTCTGG
 2641 GGGAGT GGTTTT ACCGCG CCGGCG CCGGCG CTCGTC GTCTGA TTGGCT CTCGGG GCCCAG
 2701 AAAACT GGCCTT TGCCAT TGGCTC GTTGTC GTGCAA GTTGAG TCCATC CCGCGG CCAGCG
 2761 GGGGCG GCGAGG AGGCGC TCCCAG GTTCCG GCGCTC CCGTGC GCGGCC CCGCGC AGAGTC
 2821 TGGCGG CCGGCG CTTGCG CAACGT GGCAGG AAGCGC GCGCTG GGGGCG GGGAGC GGCAGT
 2881 AGGGCT GAGCGG CTGCGG GGGGCG TGCAAG CAGGTT TCCGAC TTGAGT TGCCTC AAGAGG
 2941 GCGGTG CTGAGC CAGACC TCCATC GCGCAC TCCGGG GAGTGG AGGGAA GGAGCG AGGGCT
 3001 CAGTTG GCGTGT TTTGGA GGCAGG AAGCAC TTGCTC TCCCAA AGTCGC TCTGAG TTGTTA
 3061 TCAGTA AGGGAG CTGAG TGGAGT AGGCGG GGAGAA GCGGCG ACCCTT CTCGCG AGGGGG
 3121 GAGGGG AGTGTG GCAATA CTTTTC TGGGAG TTCTCT GCTGCC TCTGCG CTCTG AGGACC
 3181 GGCCTG GGCCTG GGAGAA TCCCTT CCCCCT CTTCCT TCGTGA TCTGCA ACTCCA GTCTTT
 3241 CCGGCG GCGCAC GTGGTA TCCGTC GACATT GATTAT TGACTA GTTATT AATAGT AATCAA
 3301 TTACGG GGTGAT TAGTTC ATAGCC CATATA TGGAGT TCCGCG TTACAT AACTTA CCGTAA
 3361 ATGGCC CCGCTG GCTGAC CGCCCA ACAGCC CCGGCG CATTGA CGTCAA TAATGA CGTATG
 3421 TTCCCA TAGTAA CGCCAA TAGGGA CTTTCC ATTGAC GTCAAT GGGTGG AGTATT TACGGT
 3481 AAACTG CCGACT TGGCAG TACATC AAGTGT ATCATA TGCCAA GTACGC CCGCTA TTGACG
 3541 TCAATG ACGGTA AATGGC CCGCTT GGCATT ATGCCG AGTACA TGAECT TATGGG ACTTTC
 3601 CTACTT GGCAGT ACATCT ACGTAT TAGTCA TCGCTA TACCA TGGTGC AGGTGA GCCCCA
 3661 GCTTCT GCTTCA CTTTCC CCACTC CCCCCT CTTCCC CAACCC GAAATT TGTATT TATTTA
 3721 TTTTIT AATTAT TTTGTG CAGCGA TGGGGG CCGGGG GGGGGG GGGGGG GCGGCG CAGGCG
 3781 GGGCGG GCGGCG GCGAGG GCGGCG GCGGCG GCGGCG GCGAGG GCGAGG GTGCGG CCGCAG CCAATC
 3841 AGAGCG GCGGCG TCCGAA AGTTTC CTTTGA TGGCGA GCGGCG GCGGCG GCGGCG CCGTAT
 3901 AAAAGC GAAGCG CCGGCG GGGGCG GAGTGC GTCGCG CTTGCC TTGCC CCGTGC CCGGCT
 3961 CCGGCG CCGCTC CCGGCG GCGGCG CCGGCG CCGGCG TCTGAC TGAACG CGTATC TCCAC AGGTGA
 4021 GCGGCG GGGAGC GCGCTT CTCCTC CCGGCT GTAATT AGCGCT TGGTTT AATGAC GCGTGG
 4081 TTTCTT TCTGTT GGTGCG GTGAAA GCGTTG AGGGGG TCCGGG AGGGGG CTTTGT GCGGGG
 4141 GAGGCG GCTGCG GGGGTT GGTGCG TGTGTT TGTGCG TGGGGA GCGGCG GCGGCG GCTGCG
 4201 CCGTGC CCGGCG GCTGTT AGCGCT GCGGCG GCGGCG CCGGGG TTTGTG CCGTCC GCAATG
 4261 TCGGCG AGGGGA GCGGCG CCGGGG CCGGTT CCGGCG GGTGCG GGGGGG GCTGCG AGGGGA
 4321 ACAAGG GCTGCG TCGGCG GTGTGT GCGTGG GGGGGT GAGTGC CCGTGC CCGGCG GGGGTT GCGGCG
 4381 TCGGCG TGCAAC CCCCC TGCACC CCGCTC CCGGAG TGTGCT AGCAGC GCGGCG CTTGCG
 4441 GTGCGG GCGTCC GTACGG GCGTGT GCGGCG GCGTGC CCGTGC CCGGCG GGGGTT GCGGCG
 4501 AGGTGG GGTGCG GCGGCG GCGGCG GCGGCG CTTGCG CCGGGG AGGGCT CCGGGG AGGGGG
 4561 GCGGCG CCGCCC GAGGCG CCGGCG GCTGTC GAGGCG CCGGGA GCCGCA GCCATT GCGTTT
 4621 TATGTT AATCGT GCGAGA GGGGCG AGGGAC TTTCTT TGTCCC AAATCT GTGCGG AGCCGA
 4681 AATCTG GGAGGC GCGGCC GCACCC CTTCTA GCGGGG GCGGGG CGAAGC GGTGCG GCGGCG
 4741 GCAGGA AGGAAA TGGGCG GGGAGG GCTTCT GTGCGT CCGGCG GCGGCC GTCCCC TTCTCC
 4801 CTCTCC AGGCTC GGGGCT GTCCCG GGGGGG ACGGCT GCGTTC GGGGGG GACGGG GCAGGG
 4861 CCGGCT TCGGCT TCTGCG GTGTGA CCGGCG GCTCTA GAGCCT CTGCTA ACCATG TTTAGT
 4921 CTTCTT TCTTCT TCTTAC AGCTCC TGGGCA ACGTGC TGGTTA TTGTGC TGTCTC ATCAT
 4981 TTGGCA AAGAAT TCGAAG CTAGCT TCGGCA TCGGCG CACCAT GACCTC TAGGAA GAAAGT
 5041 GTTGTG GAAGGT CATCAT CTTGGG GGAATC TGTGTT TGGAAA GACCTC TCTCAT GAACCA
 5101 GTATGT GAACAA GAAGTT CAGTAA CAGTAA CAAAGC CACAAAT AGGAGC GGACTT TCTGAC
 5161 CAAGGA GGTGAT GGTGGA CGACAG ACTTGT TACCAT GCAGAT CTGGGA CACAGC CGGTGA
 5221 AGAAGC GTTCCA GTCTCT TGGTGT GGCCTT CTACAG AGGTGC AGATTG CTGTGT TCTGTT
 5281 GTTGA TGTGAC TGCCCC CAACAC TTTCAA AACCTC CGAGCT CTGGAG AGACGA GTTCTT
 5341 CATCCA GGCAGC CCGGCG GGAATC CGAGAA CTTCCC TTTTGT TGTGTT GGGAAA CAAGAT
 5401 TGACCT GGAAGG CAGACA AGTGGC CACAAA GAGGGG ACAGGC TTGGTG CTACAG CAAAAA
 5461 CAACAT TCTTCA CTTCGA GACCAAG TCCCAA GGGGCG CATCAA TGTGGA GCAGGC CTTCGA
 5521 GACAAT TGCTCG GAATGC CTTTAA ACAGGA AACAGA AGTGGG ACTGTA CAATGA ATTCCC
 5581 TGAACC CATCAA ACTGGA CAAGAA TGACCG GGCCAA GGCCTC CCGAGA AAGCTG CAATGG
 5641 TTGACA ATTGTC CTCAGG TGCAGG CTGCTC ATCAGA AGGTGG TGGCTG GTGTGG CCAATG
 5701 CCGTGC CTCACA AATACC ACTGAG ATCTTT TTCCCT CTGCCA AAAATT ATGGGG ACATCA
 5761 TGAAGC CCGTGC AGCATC TGACTT CTGGCT AATAAA GGAAT TTAATT TCATTG CAATAG
 5821 TGTGTT GGAATT TTTTGT GTCTCT CACTCG GAAGGA CATATG GGGAGG CAAATC ATTAA
 5881 AACATC AGAATG AGTATT TGGTTT AGAGTT TGGCAA CATATG CCGTCA TGCTGG CTGCCA
 5941 TGAACA AAGGTT GGTAT AAAGAG GTCTAC AGTATA TGAAGC AGGCCC CTGCTG TCCATT
 6001 CTTTAT TCCATA GAAAAA CTTTGA CTGAG GTTGA TTTT TTATAT TTTGTT TTTGTT
 6061 TATTTT TTCTCT TAACAT CCGTAA AATTTT CTTTAC ATGTTT TACTAG CCAGAT TTTTCC
 6121 TCTCTT CTTGAC TACTCC CAGTCA TAGCTG TCCCTC TTCTCT TATGGA GATCAT GCATTA
 6181 GAAGAT GGGGCG GAGTCT TCTGGG CAGGCT TAAAGG CTAACC TGGTGT GTGGGG GTTGGT
 6241 CTGCG AGGGAAT TGAACA GGTGTA AAATTG GAGGGA CAAGAC TTCCCA CAGATT TTCGGT
 6301 TTTGTC GGGGAG TTTTIT AATAGG GGCAAA TAAGGA AAATGG GAGGAT AGGTAG TCATCT

Figure 2

Sequence of the final targeting vector. Homology arms are in green. KI region is in blue. Exons are bold.
 Sequence confirmed regions are underlined

6361 GGGGTT TTATGC AGCAAA ACTACA GGTGAT TATTGC TTGTGA TCCGCC TCGGAG TATTGT
6421 CCATCG AGGTAG ATTAAA GACATG CTCACC CGAGTT TTATAC TCTCCT GCTTGA GATCCT
6481 TACTAC AGTATG AAATTA CAGTGT CGCGAG TTAGAC TATGTA AGCAGA ATTTTA ATCAAT
6541 TTTAAA GAGCCC AGTACT TCATAT CCAATT CTCCTG CTCCTT CTGCAG CCTTAT CAAAAG
6601 GTATTT TAGAAC ACTCAT TTAGC CCCATT TTTATT TATTAT ACTGGC TTATCC AACCCC
6661 TAGACA GAGCAT TGGCAT TTTCCT TTTCCT GATCTT AGAAGT CTGATG ACTCAT GAAACC
6721 AGACAG ATTAGT TACATA CACCAC AAATCG AGGCTG TAGCTG GGGCCT CAACAC TGCAGT
6781 TCTTTT ATAAGT CCTTAG TACACT TTTTGT TGATCT TTGGCC TTGATC CTTAAT TTTCAG
6841 TGTTCT TCACCT CTCCTG TCAGGT GGTGTT CCACAT TTGGGC CTATTC TCAGTC CAGGGA
6901 GTTTTA CAACAA TAGATG TATTGA AAGTCC AACCTA AGCTTT AACTTT CCATC CCATGA
6961 ATGCCT CTCCTC TTTTTC TCCATT TATAAA CTGAGC TATTAA CTTAAT ATGGTT TCCAGG
7021 TGGATG TCTCCT CCCCCA ATATTA CTGTAT GTATCT TACATA TTGCCA GGCTGA TATTTT
7081 AAGACA TTAATA GGTATA TTTCAT TATTGA GCCACA TGGTAT TGATTA CTGCTT ACTAAA
7141 ATTTTG TCAATG TACACA TCTGTA AAAGGT GGTTCCT TTTTGG AATGCA AAGTTC AGGTGT
7201 TTGTTG TCTTTC CTGACC TAAGGT CTGTGT AGCTTG TATTTT TTTCTA TTAAGC AGTGCT
7261 TTCTCT TGGACT GGCTTG ACTCAT GGCATT CTACAG GTTATT CTGTGG TCAAAAT GTGATT
7321 TTGCCA AGCTTC TTGAGG ACCTAT AATTTT GCTTGA CTGTGA GCCAAA CACAAG TAAAT
7381 GATTAA GCAACA AATGTA TTGTGT AAGTCC AAGCTT AATAAA GAGAAT CTAATC TTTTGT
7441 GTGCTC TATAAT AATACT ATCCAG GGGCTG GAGAGG TGGCTC GAGATT CAAGAG CACAGA
7501 CTGTCT TTCCAG AAGTCC TGAGTT CAATTC CCAGCA ACCACA TGGTGG CTACAC ACCATC
7561 TGTAAT GGGATC TGATGC CCTCTT CTGGTG TGTTCT AAGACC ACAAGT GTATTC CATTAA
7621 AATAAA TAAATC CTCCTT CTCTCT TTTTCT TTTTAA GAGAAT CTAATC TTTTGT
7681 AGAATT TACTGA AGTAAT GAAATA CTTTGT GTTTGT TCCAAT ATGGTA GCCAAT AATCAA
7741 ATTTACT TCTTAA GCATG GAAATG TACAGA AGGAAC TAATTT TTTATT TTAAGT TAACTT
7801 TGGACA GAGGAG CCATAA CTGCAG ACTTGT GGGATA CAGAAG ACCAAT CGACAG TTTAAT
7861 GCTTTT TCTCTT ACATA AGCAAT AAAGAA ATAAAA ATTGAA CTCTCA GTATCC TATTTG
7921 TTTAAA CTGCTA GCTTTA CTAAAC TTTTGT GCTTCA TCTATA CAAAGC TGAAGG CTAAGT
7981 CTGCAG CCAATTA CTAAC ATGAAA GAAAGT AATGAT AATTTT GBAATT CAAAAA TGTAGG
8041 GCCAGA GTTTAG CCAGCC AGTGGT GGTGCT TGGCTT TATGCC TTTAAT CCAGC ACTCTG
8101 GAGGCA GAGACA GGCAGA TCTCTG AGTTTG AGCCCA GCCTGG TCTACA CATCAA GTTCTA
8161 TCTAGG ATAGCC AGGAAT ACACAG AGAAAC CTTGTT GGGGAG GGGGGC TCTGAG ATTTCA
8221 TAAAA TATAAT TGAAGC ATTCCC TAATGA GCCACT ATGGAT GTGGCT AAATCC GTCTAC
8281 CTCTCT GATGAG ATTGG GTATTA TTTTCT CTGTCT CTGCTG TTGGTT GGGTCT TTGAC
8341 ACTGTG GGTCTT CTTTAA AGCCTC CTCTCT GCACTG TGGTCT CTGTGT TGTACT TAACTT
8401 CCAATG GCTTAA ATGGCA TGGCTT TTGCC TICTAA GGGCAG CTGCTG AGATTT GCAGCC
8461 TGATTT CCAGGG TGGGGT TGGGAA ATCTTT CAAACA CTAATA TTGTCC TTTAAT TTTTCT
8521 TTTAAA AAATGG GTTATA TAATAA ACCTCA TAAAA AGTTAT GAGGAG TGAGGT GGACTA
8581 ATATTA AATGAG TCCCTC CCCTAT AAAAGA GCTATT AAGGCT TTTTGT CTTATA CTTAAC
8641 TTTTCT TTTAAA TGTGGT ATCTTT AGAACC AAGGGT CTTAGA GTTTTA GTATAC AGAAAC
8701 TGTGTC ATCGCT TAATCA GATTTT CTAGTT TCAAA CTAGAG AATCCA AATCTC TCACAG
8761 CCAAGG TCAAA TAAAGAA TTCTGT ACTTTT AATGTT AATTTG CTTACT GTGAAT ATAAAA
8821 ATGATA GCTTTT CTTGAG GCAGGG TCTCAC TATGTA TCTCTG CTTGAT CTGCAA CAAAGT
8881 ATGTAG ACTAAA GTTCTG CTTGCT TTGTCT TCTGTA ATACTA AGGTTA AATGT AGTAAT
8941 ACTTTT GGAAGT TGCAGG TCAGAT TCTTTT ATAGGG GACACA CTAAGG GAGCTT GGGTGA
9001 TAGTTG GTAAAA TGTGTT TCAAGT GATGAA AACTGT AATTAT TATCAC CGCAAC CTACTT
9061 TTTAAA AAAAAA AGCCAG GCCTGT TAGAGC ATGCTT AAGGGA TCCCTA GGAATT GCTGAG
9121 CACACA AGAGTA GTTACT TGGCAG GCTCCT GGTGAG AGCATA TTTCAA AAAACA AGGCAG
9181 ACAACC AAGAAA CTACAG TTAAGG TTACTT GTCTTT AAACCA TCTGCA TATACA CAGGGA
9241 TATTTA AATATT CAAAT AATATT TCATTC AAGTTT TCCCCC ATCAAA TTGGGA CATGGA
9301 TTCTCT CGGTGA ATAGGC AGAGTT GGAAC TAAACA AATGTT GGTTTT GTGATT TGTGAA
9361 ATTTGT TTCAGG TGATAG TTAAG CCCATG AGATAC AGAACA AAGCTG CTAATT CGAGGT
9421 CTCTGT GTTTAT ACTCAG AAGCAC TTTTGT GGGTTT CCCTGC ACTATC CTGATC ATGTGC
9481 TAGGCC TACCTT AGGCTG ATTTGT GTTCAA ATAAAC TTAAGT TTCTCT TCAGGT GATGTC

9541 ATATGA TTTTAT ATATCA AGGCCA AACATG TTATAT ATGTTA AACATT TGTACT TAATGT
9601 GAAAGT TAGGTC TTGTG GGTGTT ATTTT AATTTT CAAAA CTGAGC TAAATA AGTCAT
9661 TTTTAC ATGTCT TACATT TGGTGG AATGTT ATAATT GTGGTT TGCAGG CAAGAC TCTCTG
9721 ACCTAG TAACCC TACCTA TAGAGC ACTTTG CTGGGT CACAAG TCTAGG AGTCAA GCATTT
9781 CACCTT GAAGTT GAGAGC TTTTGT TAGTGT ATACTA GTTTAT ATGTTG GAGGAC ATGTTT
9841 ATCCAG AAGATA TCCAGG ACTATT TTGAC TGGGCT AAGGAA TTGATT CTGATT AGCACT
9901 GTTAGT GAGCAT TGAGTG GCCTTT AGGCTT GAATTG GAGTCA CTGTGA TATCTC AATAAA
9961 TGCTGG CCTTTT TAAAA AGCCCT TTCTT TTAICA CCCTGT TTCTA CATAAT TTTTGT
10021 TCAAAG AAATAC TTGTTT GGAATC CCTTTT GACAAC AATAGC ATGTTT TCAAGC CATATT
10081 TTTTCT CCTTTT TTTTCT TTTTGT TTGTTT TCGAG ACAGGG TTCTC TGATA GGCCTG
10141 GCTGCT CTGGAA CTCACT TTGTAG ACCAGG CTGGCC TCGAAC TCGAAA ATCCCG CTGCCT
10201 CTGCCT CTTAGG TGCCGG GATTAA AGGCGT GCACCA CCAGC CTGGCT AAGTTG GATATT
10261 TTGTTA TATAAC TATAAC CAATAC TAACCT CACTGG GTGGAT TTTTAA TTAGT CAGTAG
10321 TCTTAA GTGGTC TTTATT GGCCTC TCAATTA AATCTC ACTGTT CACTT AACAGA GGCTGT
10381 TGGTAC TAGTGG CACTTA AGCAAC TTTCTA CGGATA TACTAG CAGATT AAGGGT CAGGGA
10441 TAGAAA CTAGTC TAGCGT TTTGTA TACCTA CCAGCT TATAC TACCTT GTTCTG ATAGAA
10501 ATATTI CAGGAC ATCTAG AGTGTG CTATAA GGTGTA TGGTAA GCTTAT AAGGAA CTGGA
10561 AGTGGG CTAACT ACTCCA TTTCTC TGAGGG GAGAAT TAAAA TTTTGA CCAAGT GTTGT
10621 GAGCCA CTGAGA ATGGTC TCGAAA CATAAC TTTCTA AGGAAC CTCCCT AGATT CCCTCA
10681 ACTCTG CACCAC ATTTGG TCTGTC TTGAACT ATTGCT GCTGAG GATCGA TATCCA CCGCTA
10741 GGAAGT ATCCGG ACGGAT GGAATC AAGAAC CAGCCC GGGCGG TGGAGC TCGTTG ACAATT
10801 AATCAT CGGCAT AGTATA TCGGCA TAGTAT AATACG ACAAGG TGAGGA ACTAAA CCAATG
10861 GATCGG CCATGT AACAAG ATGGAT TGCACG CAGGTT CTCCGG CCGCTT GGGTGG AGAGGC
10921 TATTCG GCTATG ACTGGG CACAAC AGACAA TCGGCT GCCTGT ATGCCG CCGTGT TCCGGC
10981 TGTGAC CGCAGG GGGGCC CGGTTC TTTTGT TCAAGA CCGACC TGTGCC GTGCC TGAATG
11041 AACTCG AGGAGC AGGCAG CGCGGC TATGCT GGTGCC CACGA CGGGG TTTCTT GCGAG
11101 CTGTGC TCGAGC TGTGTA CTGAAG CGGGAA GGGAGT GGCTGC TATGTC GCGAAG TCCCG
11161 GGCAGG ATCTCC TGTAT CTACCT TGTCTC TCGCCG AGAAGT CATTGG CTGATG
11221 CAATGC GGCAGC TGCATA CGCTTG ATCCGG CTACCT GCCCAT TCGACC ACCAAG CGAAAC
11281 ATCGCA TCGAGC GAGCAC GTACTC GGAATG AAGCCG GTCTTG TGATC AGGATG ATCTGG
11341 ACGAAG AGCATC AGGGGC TCGGC CAGCCG AACTGT TCGCCA GGCCTA AGCCGC CATATC
11401 CCGAGC GCGATG ATCTCG TCGTGA CCAATG GCGATG CTTGCT TCGCCA ATATCA TGGTGG
11461 AAAATG CCGCTT TTCTGT GATTTA TCGACT GTGGCC GGTGCG GTGTGG CCGACC GCTACT
11521 AGGACA TAGCGT TGGTGA CCGGTG ATATTG CTGAAG AGCTTG GCGGCC AATGGC CTGACC
11581 GCTTCC TCGTGC TTTACG GTATCG CCGCTC CCGATT CGCAGC GCATCG CTTCTT ATGCC
11641 TCTTGT ACGAGT TCTTCT GAGGGG ATCAAT TCTCTA GAGCTC GCTGAT CAGCT CGACTG
11701 TGCCTT CTAGTT GCCAGC CATCTG TTGTTT GCCCTT CCCCC TGCCTT CTTTGA CCCTGG
11761 AAGGTG CCACTC CCACTG TCTTCT CTAAT AAAAA AGGAAA TGTGAT CGCATT GTCTGA
11821 GTAGGT GTTATT TATTCT TGGGGG GTGGGG TGGGGC AGGACA GCAAGG GGGAGG AITGGG
11881 AAGACA ATAGCA GGCATG CTGGGG ATCGGG TGGGCT CTATGG CTTCTG AGGCGC ACAGCT
11941 TTTGTT CCCTTT AGTGAG GGTAA TTTGCA GCTTGG CGTAAT CATGCT CATAGC TGTTC
12001 CTGTGT GAAATT GTTATC CGCTCA CAATTC CACACA ACATAC GAGGCC GAAGCA TAAAGT
12061 GTAAAG CTTGGG GTGCTT AATGAG TGAGCT AACTCA CATTAA TTGCGT TGCCT CACTGC
12121 CCGCTT TCCAGT CGGGAA ACCTGT CTTGCC AGCTGC ATTAAT GAATCG GCCAAC GCGCGG
12181 GGAAGC GCGGTT TGCCTA TTGGGG GCTCTT CCGCTT CTTCCG TCACT ACTCGC TGCCT
12241 CCGTCT TCGGGC TCGGCC GAGCGG TATCAG CTCACT CAAAGG CGGTAA TACGGT TATCCA
12301 CAGAAT CAGGGG ATAACG CAGGAA AGAACA TGTGAG CAAAGG GCCAGC AAAAGG CCAGGA
12361 ACCGTA AAAAGG CCGCGT TGCTGG CGTTT TCCATA GGTCTC GCCCCC CTGAGC AGCATC
12421 AAAAA ATCGAC GCTCAA GTTACA GGTGGC GAAACC CGACAG GACTAT AAGAT ACCAGG
12481 CGTTCT CCGCTG GAAGCT CCGTGC TGCCTC CTCTGT TTTGCA CCCTGC CGCTTA CCGGAT
12541 ACCTGT CCGCTT TTTCTC CTTCCG GAAGCG TGGCGC TTTCTC ATAGCT CAGCT GTAGGT
12601 ATCTCA GTTCGG TGTAGG TCGTTC GCTCCA AGCTGG GCTGTG TGCAGC AACCCC CCGTTC
12661 AGCCCG ACCGCT GCGCT TATCCG GTAAC TATGTC TTAGT CCAACC CGGTAA GACAGC

Figure 3

Sequence of the final targeting vector. Homology arms are in green. KI region is in blue. Exons are blond. Sequence confirmed regions are underlined.

Figure 4

The identification of PCR and southern blot. A. The strategy of PCR. B. The results of PCR. C. The strategy of southern blot analysis. D. The results of southern blot.

Figure 5

Sequencing confirmation. A. The strategy of sequencing confirmation. B. The results of sequencing.