Design CRISPR-Cas9 Guide RNA and Donor Oligo to Introduce Pathogenic Mutation into Mice

Xijing Liao[1*], Xueyan Yang[2]

[1] YK Pao School, Shanghai 201620, China

[2] School of Life Sciences, Fudan University, Shanghai 200433, China

Email: 450226983@qq.com

Abstract: CRISPR-Ca9 system is a newly developed gene-editing technology, which is widely used in biology and medical research. In this project, we want to knock-in a mutation found in a human pedigree into mice through CRISPR-Cas9 technology to validate its pathogenic effect. We download corresponding mice genomic sequences and design guide RNA and donor oligo sequences according to CRISPR-Cas9 target principles. Following experiments confirm that this set of sequences is effective in mice cell line.

Keywords: CRISPR-Ca9, RNA, DNA sequences

1. Introduction

CRISPR-Cas9 is a new gene-editing technology, which originates from microbial adaptive immune system that uses RNA-guided nucleases to cleave foreign genetic elements [1]. CRISPRs are specialized stretches of DNA that encodes the guide RNAs and a required auxiliary trans-activating crRNA (tracrRNA) that facilitates the processing of the crRNA array into discrete units. CRISPR can guide and localize Cas9 to specific genomic region. The protein Cas9 is an enzyme that acts like a pair of molecular scissors to cut DNA strands. CRISPR-Cas9 promotes genome editing by stimulating double strand breaks (DSB) at a target genomic locus. The target locus typically undergoes different pathways for DNA damage repair: the error-prone NHEJ (non-homologous end-joining) or the high-fidelity HDR (homology-directed repair) pathway. Usually, NHEJ will lead to gene knock-out and HR will result in gene knock-in [2]. Comparing to other nuclease technology such as ZFNs[3] and TALENs[4], CRISPR-Cas9 offer several potential advantages including the ease of customization, higher targeting efficiency and the ability to facilitate multiple genome editing. This editing tool has been widely applied in a variety of fields including basic biological research, development of biotechnological products and treatment of diseases. Two founders of CRISPR-Cas9, Emmanuelle Charpentier and Jennifer were awarded by the Nobel Prize in chemistry in 2020.

In our previous study, we find an autosomal dominant pedigree, which contains 4 generations and 15 affected individuals (Figure 1A). We use linkage analysis to locate candidate chromosome region on 6q27 with LOD > 2.0. Among this region we find TBXT gene which is relate to vertebrate development. Then we discover that all affected members in this pedigree harbor a heterozygous missense change in the highly conserved T-box domain of TBXT gene (NM_003181, c.596A>G, p.Gln199Arg) (Figure 1B). The 199Q site on TBXT is highly conserved among different species (Figure 1C). To validate the pathogenic effect of TBXT Q199R mutation, we plan to design the guide RNA and donor oligo sequences and construct knock-in mice via CRISPR-Cas9 technology.
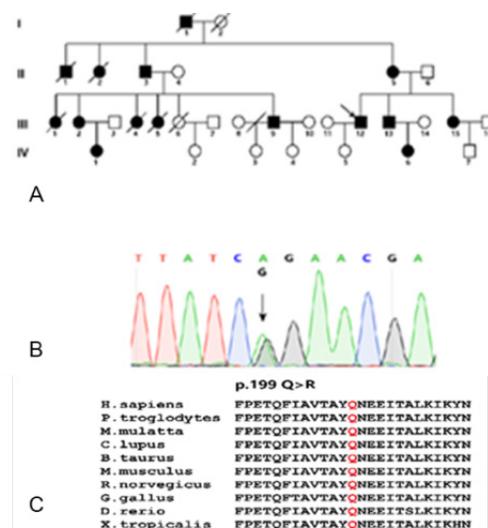

Figure 1. TBXT p.Q199R mutation was found in a pedigree

Note: A) Human pedigree. B) Genomic DNA sequence of the proband. C) Alignment of TBXT ortholog protein sequences. p.Q199 residues are highlighted in red.

2. Results

2.1 Design guide RNA sequence for CRISPR-Cas9 targeting

First, we download the mouse TBXT mRNA (NM_009309.2) and corresponding genomic sequence (chr.17) from NCBI database (www.ncbi.nlm.nih.gov) and distinguish the exon region and intron region. Then we find out the p.Q199 site according to its protein sequence (NP_114394.1). Here is the mouse genomic sequence containing p.Q199 region (Figure 2).

Figure 2. Genomic sequence of mouse TBXT gene contains Q199 site

Note: Red uppercase letters are exon4 sequence. Yellow highlight region is Q199 site. Black lowercase letters are intron sequence.

Second we screen this region for the guide RNA sequences fulfill several principles: a) The length of sequence is 20-nt; b) There is 5′-NGG sequence downstream the guide sequence; c) this sequence is adjacent to p.Q199 site (CAG). We find lots of candidate sequences meet these requirements. Because we want to do knock-in instead of knock-out in mice, we prefer the guide RNA sequence near by the Q199 site. It is better that the Cas9 nuclease only cut the original sequence and doesn't cut the replaced sequence by donor oligo so that it can raise the knock-in efficiency. Thus we only pick up 3 sequences around Q199 site as shown in Figure 3 and Table 1.
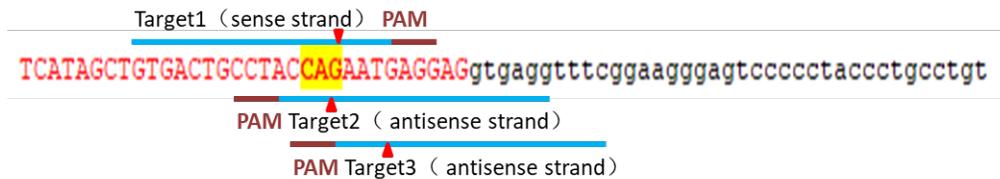
Figure 3. Candidate guide RNA sequences

Note: Red uppercase letters are exon4 sequence. Yellow highlight region is p.Q199 site. Black lowercase letters are intron sequence. Blue line is target sequence. Brown line is PAM sequence. Red arrowhead is Cas9 nuclease cut site.

Third, we blast these sequences in mouse genome to pick up the one with the minimization of off-target activity. Here is the list of these sequences and their blast results. According to the blast result, target1 has the highest possibility to off-target because the highest similarity other than Tbxt itself is 95% (only 1 base differs from Tbxt 19). While target3 has the lowest possibility to off-target for the highest similarity other than Tbxt itself is 85% (3 bases differ from chromosome 5). Considering that we can destroy the PAM sequence of target 3 when we change CAG to arginine codon, target 3 is the optimal choice for knock-in.

Table 1. The blast results of three sequences

Name	Sequence	Blast result
Target 1	GTGACTGCCTACCAGAATGA	Distribution of the top 113 Blast Hits on 100 subject sequences PREDICTED: Mus musculus T-box 19 (Tbx19), transcript variant X1, mRNA Sequence ID: XM_006497043.3 Length: 2429 Number of Matches: 1 Range 1: 533 to 552 GenBank Graphics ▼ Next Match ▲ Score: 32.2 bits(16) Expect: 3.4 Identities: 19/20(95%) Gaps: 0/20(0%) Strand: Plus/Plus Query 1 GTGACTGCCTACCAGAATGA 20 |||||||||||| |||||||| Sbjct 533 GTGACTGCCTATCAGAATGA 552
Target 2	aaacctcacCTCCTCATTCT	Distribution of the top 134 Blast Hits on 100 subject sequences Mus musculus BAC clone RP23-84E16 from 7, complete sequence Sequence ID: AC127598.3 Length: 186752 Number of Matches: 1 Range 1: 127461 to 127478 GenBank Graphics ▼ Next Match Score: 36.2 bits(18) Expect: 0.22 Identities: 18/18(100%) Gaps: 0/18(0%) Strand: Plus/Plus Query 3 ACCTCACCTCCTCATTCT 20 |||||||||||||||||| Sbjct 127461 ACCTCACCTCCTCATTCT 127478
Target 3	ctcacCTCCTCATTCTGGTA	Distribution of the top 150 Blast Hits on 100 subject sequences Mus musculus chromosome 5, clone RP23-123P10, complete sequence Sequence ID: AC163094.5 Length: 208850 Number of Matches: 1 Range 1: 79393 to 79409 GenBank Graphics ▼ Next Match Score: 34.2 bits(17) Expect: 0.85 Identities: 17/17(100%) Gaps: 0/17(0%) Strand: Plus/Plus Query 3 CACCTCCTCATTCTGGT 19 ||||||||||||||||| Sbjct 79393 CACCTCCTCATTCTGGT 79409

2.2 Design donor DNA sequence to change Tbxt p.Q199 to p.R199.

Because we want to introduce a mutation into mouse genome, we need a repair template to leverage the HDR pathway, which allows high fidelity and precise editing. Although plasmid or single-stranded DNA oligonucleotides (ssODNs) can be used as repair template, ssODNs have been used in place of targeting plasmids for minor modifications within a defined locus. To achieve high HDR efficiencies, ssODNs must contain at least 40bp flanking sequences on each side which are homologous to the target region. Both the sense and antisense sequence of the target locus can be used as ssODNs.

In our case, we first determine the site we want to change and choose 40bp on each site. Thus the original sequence is shown here.

CTGCTTTCCCGAGACCCAGTTCATAGCTGTGACTGCCTAC**CAG**AATGAGGAGgtgaggtttcggaagggagtccccctaccct

Figure 4. The original sequence

Then we check codons for arginine and find that there are 6 codons encoding arginine including CGA\CGU\CGT\CGG\AGA\AGG. Because the first C in CAG also contributes to the PAM sequence, we should change it to A to avoid future cutting by Cas9 nuclease after homologous recombination. Thus we finally change CAG to AGG to make the knock-in. The final ssODN sequence is shown here.

CTGCTTTCCCGAGACCCAGTTCATAGCTGTGACTGCCTAC**AGG**AATGAGGAGgtgaggtttcggaagggagtccccctaccct

Figure5. The final ssODN sequence

3. Conclusion

In this project, I review the development of gene editing technology. Then I apply the design principles of CRISPR-Cas9 into the specific case and pick up a set of guide RNA and ssODN to construct knock-in mouse. Following experiments confirm that this set works very well in mice (data not shown). Through this procedure, I learn and practice various biological study skills including literature retrieval, database searching, DNA/protein sequence analysis and experimental design.

References

[1] Ishino, Y., M. Krupovic and P. Forterre, History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology. *Journal of Bacteriology*, 2018. 200(7): p. e00580-17.
[2] Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc-finger nucleases. *Nat. Rev. Genet*. 11, 636–646 (2010).
[3] Ran, F.A., et al., Genome engineering using the CRISPR-Cas9 system. Nature Protocols, 2013. 8(11): p. 2281-2308.
[4] Porteus, M.H. and D. Carroll, Gene targeting using zinc finger nucleases. *Nature Biotechnology*, 2005. 23(8): p. 967-973.
[5] Sommer, D., et al., TALEN-mediated genome engineering to generate targeted mice. *Chromosome Research*, 2015. 23(1): p. 43-55.
[6] CRISPR-CAS9, TALENS and ZFNS - the battle in gene editing. Available from: https://www.ptglab.com/news/blog/crispr-cas9-talens-and-zfns-the-battle-in-gene-editing/
[7] Press release: The Nobel Prize in Chemistry 2020. Nobel Foundation. Retrieved 7 October 2020. Available from: https://www.nobelprize.org/prizes/chemistry/2020/press-release/