Selection of Grna for Genomic Editing of the Bovine Leucosis Virus Susceptible Alleles of the 2 Exon of the Bola-DRB3 Gene by CRISPR/Cas9

Anastasia S. METLEVA

Ph.D. (in Veterinary Sciences)

Associate Professor

Department of Zootechnics, Faculty of Zootechnics

Kuzbass State Agricultural Academy

9a, Central Str., Kemerovo, 650510, Russia

Konstantin V. BESPOMESTNYKH

Ph.D. (in Technical Sciences)

Head of Laboratory

Department of Zootechnics, Faculty of Zootechnics

Kuzbass State Agricultural Academy

9a, Central Str., Kemerovo, 650510, Russia

Abstract

A great spread of the cattle leucosis virus among the livestock of the Russian Federation and the lack of reliable methods of its treatment and prevention led to the creation of animals resistant to the leucosis virus by genetic engineering. In order to create high-tribal cattle resistant to leucosis virus, the genomic editing method CRISPR/Cas9 was used. The work is devoted to the first and main stages of the genomic editing method - the choice of the method of introducing the target modification; the choice of the method of repair of the selected site, the selection of DNA sequencing, optimally suitable for the occurrence of the target mutation; analysis of off-target sites and the choice of guiding gRNA. As a result of the work, using molecular genetic programs, an allele sensitive to the virus was selected, non-targeted sites were identified and analyzed, and gRNA was selected.

Keywords: cattle leucosis; genomic editing; CRISPR/Cas9; guiding gRNA; knockout gene.

Introduction

Genetic engineering is the most promising and developing branch of biological science, able to get ahead of the traditional breeding process. The creation of new animal lines allows getting animals with valuable breeding qualities that are resistant to various diseases at genetic level.

As of March 2020, 67 constituent entities of the Russian Federation including Kemerovo region are unsuccessful in cattle leucosis, which is confirmed by official information from the Federal Service for Veterinary and Phytosanitary Surveillance <https://cerberus.vetrif.ru/cerberus/regionalization/pub>.

Taking into account the lack of vaccines or other means of treatment and prevention, as well as the ambiguous position of regulatory acts regulating the circulation of raw milk from cows with leucosis virus, the creation of a line of animals resistant to the cattle leucosis virus may become a trend in biotechnology aimed at to fight this infection.

To obtain the target modification to the virus-sensitive allele with the aim of editing it into a stable one, the genetic engineering method - CRISPR/Cas9 was chosen. Using molecular genetic electronic resources, a gRNA that optimally satisfies the requirements was selected and analyzes off-target sites with the largest number of nucleotide mismatches of gRNA was conducted.

Methods

In order to accomplish the tasks, there were used online tools that determine the site for introducing gaps in the genomic sequence of interest and identifying suitable target sites.

Bovine leucosis is a viral disease characterized by malignant proliferation of lymphatic tissue. The causative agent of this disease is the bovine leucosis virus (BLV), which belongs to the family of RNA-containing viruses of the family Retroviridae, genus Deltaretrovirus (Polat et al., 2017).

The development of leucosis in the body is associated with many factors, including the state of the animal's immune system, as well as its predisposition to the disease at the genetic level (Kovalyuk, 2008; van Eijk et al., 1992). The primary gene, BoLA-DRB3, the main histocompatibility complex containing 6 exons, namely, exon-2 of the BoLA-DRB3 gene, consisting of 284 bp, is responsible for the body's primary immune response to the introduction of BLV and the recognition of the antigen and the formation of an immune response system for BLV.

Numerous scientists have found that the different exon-2 alleles of this gene are associated with resistance or resistance to BLV (Kovalyuk, 2008; Ernst et al., 1997; Juliarena et al., 2008; Mirsky, 1998; Nassiry et al., 2005).

For example, analysis of amino acid sequences showed that alleles (*11, *23, *28) responsible for coding the amino acid sequence of Glu-Arg did not occur in animals infected with BLV. And animals infected with BLV do not obtain a single allele that encodes the amino acid sequence of ER (glutamic acid and arginine). But animals with signs of the leucosis virus have the *16 allele as the most common. Sick individuals also have alleles *8 and *22. Animals suffering from leucosis are characterized by the presence of the VDTN sequence (valine - aspartic acid - threonine -

asparagine) at position 75-78 (alleles *16 and *8), as well as the VDTV sequence (valine - aspartic acid - threonine - valine) (*22 allele) These alleles are associated with the alleles of the BoLA-DRB3*1C gene, for which a correlation with predisposition to BLV was previously shown (Ran et al., 2015).

Numerous works of various research groups around the world on various cattle breeds give grounds to consider it reliable to use *11, *23 and *28 alleles as markers of cattle resistance to leucosis. Alleles *8, *16, *22 and *24 are considered to be alleles of sensitivity to the virus (Kozlov, 2016).

Considering the fact that the resistance and susceptibility of cattle to BLV is genetically determined, it is possible to develop a technology for creating cattle resistant to BLV by genetic engineering methods, in particular using the CRISPR/Cas9 genome editing technology. Such an approach will allow not only getting the cattle genetically resistant to leucosis during two or three regenerations, but also avoiding a long narrowly targeted breeding process that helps to reduce genetic diversity in the population, which will negatively affect productivity and resistance to other infections. In addition, there is a negative correlation between milk production and resistance to leucosis virus. Thus, the technology of creating cattle populations resistant to the leucosis virus by molecular genetics and genomic editing has higher potential than traditional breeding methods.

The CRISPR/Cas9 system is an advanced tool in modern genetic engineering, one of the applications of which is the correction of mutations that cause hereditary diseases, both in vitro and in vivo (Menzorov et al., 2016; Ran et al., 2015; Cong et al., 2013; Cox et al., 2015).

The CRISPR/Cas9 system consists of two main parts. The first is a protein, Cas9 nuclease, which is capable of introducing a double-stranded break in a DNA molecule. The second is a small RNA molecule with a size of about 120 nucleotides - a chimeric guide RNA (gRNA) (Menzorov et al., 2016).

Editing the genome using the CRISPR/Cas9 system has a number of difficulties that arise during the work, including selection of the sequence of the guiding gRNA, analysis of its specificity, and selection of sequences with the least number of off-target sites, because there is a high probability of manifestation of inappropriate effects (Juliarena et al., 2008). Non-target sites differ from the target sequence by a few nucleotides, which lead to undesirable mutations and chromosomal rearrangements. To solve this problem, it is necessary to select the appropriate tools to detect potential non-targeted sites. These tools carry out an experimental evaluation of non-target genome modifications for each specific gRNA and calculate non-target sites.

In order to efficiently transcribe gRNA, it is necessary to observe some rules in the process of selecting it:

1. The optimal sites for introducing double-stranded DNA breaks with the aim of knocking out the gene is a portion of the exon region of DNA located at a distance of about 150-300 bp. after the start codon (Zakiyan et al., 2018).
2. The first nucleotide located at the 5'-end of the gRNA (from which the synthesis of the RNA molecule will begin) should be guanine. Therefore, it is necessary to select DNA sequences of the type G(N)₁₉ NGG.
3. It is necessary that the 17th nucleotide of the protospacer directly adjacent to the rupture site, and most affecting the outcome of the repair, to be thymine or adenine. In this case, insertion deletion with a shift of the reading frame is more likely to occur (Bae et al., 2014).
4. It is necessary to ensure the elimination of the occurrence of non-target mutations with the effective selection of target ones.

Without the use of a computer and appropriate software, such an analysis is impossible. Moreover, many programs rank selected protospacers according to their suitability for editing and eliminating inappropriate mutations, giving them a quantitative assessment (Ernst et al., 1997).

Results

To edit the cattle genome in order to create lines resistant to the leucosis virus, the method of modifying the sensitive allele (*16) of the exon-2 of the BoLA-DBR3 gene using double-stranded rupture followed by non-homologous repair was chosen, because non-homologous restoration of damage generated by Cas9 contributes to the creation of a zero allele ("knockout") in the edited gene, which affects the occurrence of errors in the form of insertions and deletions. Such errors generated during non-homologous repair are usually small (1-10 bp). In non-homologous repair of the gap, the site identical to the amino acid chain of the reverse transcriptase of the leukemia virus loses its identity due to a change in the nucleotide sequence, which entails a shift in the reading frame during transcription and to a change in the amino acid composition during protein biosynthesis or to the formation of a stop codon, at which it stops transcription, in this case, the protein is shortened and not functional (Menzorov et al., 2016).

In order to create cattle lines resistant to leucosis virus using genome editing technologies using the CRISPR/Cas9 system, the BoLA-DBR3 histocompatibility complex gene located on the 23rd chromosome of the cattle genome is optimal, which is confirmed by the NCBI molecular biological resources database (<https://www.ncbi.nlm.nih.gov/>) (Figure 1).

Figure 1. Location of the main histocompatibility complex BoLA-DRB3 on the 23d chromosome

At the first stage, which consists in identifying the target sequence, the nucleotide sequence of the exon-2 16th allele of the BoLA-DRB3 histocompatibility complex gene (Figure 2) containing the (VDTY) motif responsible for BLV susceptibility was analyzed in the international NCBI genetic database (Figure 3).

Figure 2. The nucleotide sequence of the exon-2 16th allele gene of the histocompatibility complex BoLA-DRB3

Figure 3. Amino acid sequence of the exon-2 16th allele gene histocompatibility complex BoLA-DRB3

The analysis found that in the course of the experiment double stranded break must be introduced into the target gene site BoLA-DRB3, comprising a sequence of nucleotides: «GGAGTATTCTACGAGCGAGTGTTCATTTCTTCAACGGGACCGAGCGGGGTGCGGTACCTG GACAGATACTTCCATAATGGAGAAGAGTTCGTGCGCTTCGACAGCGACTGGGGCGAGT ACCGGGCGGTGACCGAGCTAGGGCGGCGGTCCGCCGAGCAGTTGAACGGCCAGAAGG ACACCCTGGAGCGGGAGCGGGCCTATGTGGACACGTACTGCAGACACA ACTACGGGGT CGTTGA», encoding the amino acid sequence Val-Asp-Thr-Tir (VDTY) (Figure 3), which is associated with susceptibility to leucosis.

The next stage of work was the selection of gRNA, using one of the CHOPCHOP web tools (<https://chopchop.cbu.uib.no/>) to select the CRISPR/Cas9 target sites. The selection included the following steps:

1. The introduction of the nucleotide sequence obtained as a result of the analysis of molecular biological programs;
2. The choice of an organism from the proposed list. In our case, the selected organism is *Bos Taurus*;
3. Indications of the system with a help of which it is planned to carry out editing. The CRISPR/Cas9 system was chosen (Figure. 4).

Figure 4. Selection of gRNA for editing the Bola-DRB3 gene using the molecular biological program CHOPCHOP (<https://chopchop.cbu.uib.no/>)

When working with a computer program for selecting gRNAs, the results of DNA sequences homologous to gRNAs are presented in the form of a graphic image of the search area with designated sequences of sequences that are complementary to gRNAs and are supplemented by a table with all possible sequences of sequences.

Priority in areas for gRNA is indicated with colors (green - optimal, red - worse) (Figure 4) (Zakiyan et al., 2018).

Based on the recommended nuances in the selection of gRNA, we selected the sequence number 22, which represents the following nucleotide sequence - GGAGCGGGAGCGGGCCTATGTGG, in which at the 5'-end the guide should have 20 nucleotides adjacent to the PAM site: GGAGCGGGGAGCGGGUCCA (Figure 5).

Home	Instructions	Scoring	About	Updates	Submissions	Contact	FAQ
------	--------------	---------	-------	---------	-------------	---------	-----

Target: **fastalInput.fa**
Rank: **22**
Target sequence: **GGAGCGGGAGCGGGCCTATGTGG**

Off-targets		
Location	Number of mismatches	Sequence (including mismatches)
*:0	0	iiiiiiiiiiiiiiii
chr1:11033932	3	CCACATAGtCCtGCTCCCTCTCC
chr11:104906926	3	GGAGCaGGAGCGGGCCcAgGAGG
chr19:36191706	3	CCCCATgGGCCaGcACCCGCTCC
chr2:129991935	3	GcAGgGGAGgGGGCTATGGG
chr3:29272639	3	CCTCATAGcCCCTCTCCCTCTCC
chr3:120850279	3	CCCCAcAGGCCaGCTCCCGCTGc

Shen et al. 2018 predictions of repair profile - statistics		
Reference sequence	GGCGCGGTCCGCCGAGCAGTTGAACGGCCAGAAGGACACCTGGAGCGGGAGCGGGCTA<>TGTGGACACGTACTGCAGACACAACACTACGGGTCGTTGA	
Frameshift frequency		68.68
Precision score		0.42
Frame +0 frequency		31.32
Frame +1 frequency		37.50
Frame +2 frequency		31.19
1-bp ins frequency		16.29
Highest del frequency		8.10

Figure 5. Nonspecific sites of selected gRNA

The selected DNA section sequence meets the basic requirements for the guide selection:

1. The first nucleotide located at the 5'-end of the DNA is guanine and corresponds to the formula - G(N)19 NGG.
2. Thymine (T) is the 17th nucleotides of the protospacer directly adjacent to the break, which is likely to affect the insertion deletion with a shift in the reading frame.
3. The indicated site has the site: TGG
4. The indicated sequence has off-target sites in the amount of 6 non-target sites with 3 nucleotide sequence inconsistencies of the selected sequence.

Discussion

The selected allele for editing corresponds to literature data: it is susceptible to the cattle leucosis virus and contains the amino acid motif VDTY, which is found in animals with BLV. An experiment on introducing a double-stranded break by knockout followed by non-homologous repair of the selected site using the CRISPR/Cas9 method is capable of moving the reading frame with a high probability, followed by the production of a non-functional protein.

The program for the selection of gRNA on the selected sequence does not offer DNA sections that are optimally suited for the introduction of the genetic construct: satisfying all the requirements; with the absence of off-target sites and visualized by the program in green.

The program offers sites highlighted in orange with a minimum number of non-specific sites (no more than 3). When analyzing the areas highlighted in orange, it was found that they have non-specific sites that are absolutely identical to the selected sequence, which will entail undesirable

mutations and such sites are not suitable for editing. In addition, these genome sections did not meet the requirements, compliance with which is highly desirable when selecting gRNA.

There was selected gRNA under serial number 22. The indicated region begins with guanine; the 17th nucleotide is thymine; all non-specific sites contain 3 nucleotide substitutions, which reduce the possibility of transcription of gRNA to off-target sites.

Conclusion

As a result of the work performed, a genomic region was selected for introducing a double-stranded gap with subsequent non-homologous repair of the exon-2 16th allele of the Bola-DRB3 gene. The selection of gRNA was carried out. To minimize the variants of non-target mutations, molecular biological information resources were used: NCBI, CHOPCHOP.

Acknowledgments

For the joint work, we express our gratitude to the colleagues of FSBEI HE St. Petersburg State Academy of Veterinary Medicine.

References

- Bae, S., Kweon, J., Kim, H.S., Kim J.S. (2014). Microhomology-based choice of Cas9 nuclease target sites. *Nature Methods*, 11(7), 705-706. <https://doi.org/10.1038/nmeth.3015>
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F. (2013). Multiplex genome engineering using CrisPr/Cas systems. *Science*, 339(6121), 819-823. <https://doi.org/10.1126/science.1231143>
- Cox, D.B.T., Platt, R.J., Zhang, F. (2015). Therapeutic genome editing: prospects and challenges. *Nature medicine*, 21(2), 121-131. <https://doi.org/10.1038/nm.3793>
- Ernst, L.K., Sulimova, G.E., Orlova, A.R., Udina, I.G., Pavlenko, S.P. (1997). Features of BoLA-DRB3 antigen distribution in black-motley cattle in connection with the association with leukemia. *Genetics*, 1, 84-95.
- Juliarena, M.A., Poli, M., Sala, L., Ceriani, C., Gutierrez, S., Dolcini, G., Rodríguez, E.M., Mariño, B., Rodríguez-Dubra, C., Esteban, E. N. (2008). Association of BLV infection profiles with alleles of the BoLA-DRB3.2 gene. *Animal Genetics*, 39(4), 432-438. <https://doi.org/10.1111/j.1365-2052.2008.01750.x>
- Kovalyuk, N.V. (2008). Molecular-biological methods for treating cattle from leukemia. *Veterinary medicine*, 2, 22-26. (in Russia)

- Kovalyuk, N.V. (2008). *Molecular-genetic aspects in breeding and early diagnosis of cattle leukemia*. Ph.D. dissertation. Krasnodar. (in Russian)
- Kozlov, A.L. (2016). *Polymorphism of the Bola-DRB3 gene as a marker for assessing genetic diversity and resistance to leukemia viruses in dairy cattle of the Bryansk region*. Ph.D. dissertation. Stavropol. (in Russian)
- Menzorov, A.G., Lukyanchikova, V.A., Korablev, A.N., Serova, I.A., Fishman, V.S. (2016). A practical guide to editing genomes of the CRISPR/Cas9 system. *The Vavilov's Journal of Genetics and Breeding*, 20(6), 930-944. (in Russian)
- Mirsky, M.L. (1998). Reduced bovine leukaemia virus proviral load in genetically resistant cattle. *Animal Genetics*, 29, 145-252.
- Nassiry, M.R., Shahroodi, F.E., Mosafer, J., Mohammadi, A., Manshad, E., Ghazanfari, S., Mohammad Abadi, M.R., Sulimova, G.E. (2005). Analysis and frequency of bovine lymphocyte antigen (BoLA-DRB3) alleles in Iranian Holstein cattle. *Genetika*, 41(6), 817-822.
- Nemudryi, A.A., Valetdinova, K.R., Medvedev, S.P., Zakian, S.M. (2014). TALEN and CRISPR/Cas Genome Editing Systems: Tools of Discovery. *Acta naturae*, 6(3), 20-42.
- Polat, M., Takeshima, S.-N., Aida, Y. (2017). Epidemiology and genetic diversity of bovine leukemia virus. *Virology Journal*, 14(1), 209. <https://doi.org/10.1186/s12985-017-0876-4>
- Ran, F.A., Cong, L., Yan, W.X., Scott, D.A., Gootenberg, J.S., Kriz, A.J., Zetsche, B., Shalem, O., Wu, X., Makarova, K.S., Koonin, E.V., Sharp, P.A., Zhang, F. (2015). In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, 520(7546), 186-191. <https://doi.org/10.1038/nature14299>
- van Eijk, M.J., Stewart-Haynes, J.A., Lewin, H. A. (1992). Extensive polymorphism of the BoLA-DRB3 gene distinguished by PCR-RFLP. *Animal Genetics*, 23(6), 483-496. <https://doi.org/10.1111/j.1365-2052.1992.tb00168.x>
- Zakiyan, S.M., Medvedev, S.P., Dementieva, E.V., Pokushalov, E.A., Vlasov, V.V. (2018). *Editing genes and genomes* (Vol. 3). Publishing House of the SB RAS. (in Russian)