A Machine Learning approach for the identification of CRISPR/Cas9 nuclease off-target for the treatment of Hemophilia

Trishang Udhwani[1], Anuraj Nayarisseri[1,2,*]

1. *In silico* Research Laboratory, Eminent Biosciences, Mahalakshmi Nagar, Indore – 452010, Madhya Pradesh, India.

2. Bioinformatics Research Laboratory, LeGene Biosciences Pvt Ltd., Indore - 452010, Madhya Pradesh, India.

* Corresponding authors:

Dr. Anuraj Nayarisseri
Principal Scientist and Director
In silico Research Laboratory, Eminent Biosciences, Mahalakshmi Nagar, Indore – 452010, Madhya Pradesh, India.
Email:anuraj@eminentbio.com
Tel: +91 9752295342

Hemophilia can be defined as a genetic disorder in which the body loses its capability to clot blood, and hence can't stop blood flow. It is an X- linked recessive disease, hence mostly seen in males, with its severity significantly reduced in females. In India, hemophilia has an occurrence of 1 per 10,000 births, which generally progresses to a chronic disability or premature death in subjects left untreated or provided with suboptimal treatment, a case prevalent in India [1]. The two major types of this condition are hemophilia A or factor VIII deficiency and hemophilia B or factor IX deficiency. In addition, hemophilia C is a rare category in which inhibitory antibodies develop which show high affinity to procoagulants, thus neutralizing the effect of a coagulation factor. Inhibitors are much less common in patients with hemophilia B than in those with hemophilia A. The F8 gene present on the X chromosome is responsible for guiding the production of coagulation factor VIII, essential for forming blood clots. In hemophilia A, mutation results from two gross (140 kbp or 600 kbp) chromosomal inversions that involve introns 1 and 22, respectively. Similarly, the F9 gene present on the X chromosome mutates through several different mechanisms to give rise to hemophilia B condition. The conditions associated with hemophilia Leyden, ribosome readthrough of nonsense mutation and apparently 'silent' changes that do not alter amino acids are the major mutations studied. It has been observed that reconstitution with 1–2% of the clotting factor helps uplift the quality of life, while 5–20% reconstitution is required to ameliorate the genetic disorder. Gene-specific genome editing is preferred over random integration of expression cassettes as this helps avoid genotoxicity and achieve the required physiological levels of expression. Advances in genome

engineering based on CRISPR- associated RNA- guided endonuclease Cas9 are empowering the guidance of the said endonuclease to target locations by a short RNA search string [2]. It requires a programmable sequence-specific RNA to direct it and introduce cleavage and double-stranded breaks at the target site. In case of hemophilia A, induced pluripotent stem cells (iPSCs) can be derived from patients with inversion genotypes with an aim to revert these chromosomal conditions to the corrected state with the assistance of CRISPR- Cas9 nucleases [3]. The endothelial cells from the corrected iPSCs can be checked for expression of F8 gene and the production of factor VIII. Likewise, in case of hemophilia B, delivery of naked Cas9-sgRNA plasmid and donor DNA, aiming to recover the mutation has shown a detectable gene correction (>1%) in F9 alleles of hepatocytes [4]. To construct the related plasmids, an AAVS1-Cas9-sgRNA plasmid is designed to cut the AAVSI locus in human. Subsequently, two donor plasmids are designed to insert GFP and F9 cDNA into the designated AAVS1 locus. Whole genome sequencing (WGS) is used in combination with this editing method to identify off-target mutations, to ensure that editing takes place at the desired site. The technique offers several benefits over the popularly used Adeno- associated viral (AAV) vectors such as precision, decreased insertional oncogenesis and control through an endogenous promoter [5]. The CRISPR/Cas9- mediated genome editing with an AAV8 vector has been put to use to provide an adjustable path to induce double-strand breaks at the target genes in hepatocytes [6]. The foremost need for CRISPR-Cas9 is the identification of targets that have undergone a mutation, which has led to the development of the said condition. Although a few targets are known, none of the target mutation has been capable to render a 5-20 percent of reconstitution that is required for the elimination of the disorder. Hence, there is a need to find novel targets for the CRISPR-Cas9 system, which in turn requires the assistance of computational tools. The aim of this study is to identify positive CRISPR-Cas9 targets which would help in better and more accurate treatment of the disorder with computational biology facilitating the research. The study provide targets which possess minimum off-target mutations, providing maximum reconstitution for hemophilia.

References

[1] Kar, A., Phadnis, S., Dharmarajan, S., & Nakade, J. (2014). Epidemiology & social costs of haemophilia in India. *The Indian journal of medical research*, *140*(1), 19.

[2] Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, *157*(6), 1262-1278.

[3] Park, C. Y., Kim, D. H., Son, J. S., Sung, J. J., Lee, J., Bae, S., ... & Kim, J. S. (2015). Functional correction of large factor VIII gene chromosomal inversions in hemophilia A patient-derived iPSCs using CRISPR-Cas9. *Cell stem cell*, *17*(2), 213-220.

[4] Huai, C., Jia, C., Sun, R., Xu, P., Min, T., Wang, Q., ... & Lu, D. (2017). CRISPR/Cas9-mediated somatic and germline gene correction to restore hemostasis in hemophilia B mice. *Human genetics*, *136*(7), 875-883.

[5] Doshi, B. S., & Arruda, V. R. (2018). Gene therapy for hemophilia: what does the future hold?.*Therapeutic advances in hematology*, *9*(9), 273-293.

[6] Ohmori, T., Nagao, Y., Mizukami, H., Sakata, A., Muramatsu, S. I., Ozawa, K., ... & Sakata, Y. (2017). CRISPR/Cas9-mediated genome editing via postnatal administration of AAV vector cures haemophilia B mice. *Scientific reports*, *7*(1), 4159.