

HKUMed researchers discover more efficient CRISPR-Cas9 variants for gene therapy applications

Reviewed by Emily Henderson, B.Sc.

Jun 30 2022

A research team from the LKS Faculty of Medicine, The University of Hong Kong (HKUMed) discovered more efficient CRISPR-Cas9 variants that could be useful for gene therapy applications. By establishing a new pipeline methodology that implements machine learning on high-throughput screening to accurately predict the activity of protein variants, the team expands the capacity to analyze up to 20 times more variants at once without the need for acquiring additional experimental data, which vastly accelerates the speed in protein engineering. The research team has successfully applied the pipeline in several Cas9 optimizations and engineered new *Staphylococcus aureus* Cas9 (SaCas9) variants with enhanced gene editing efficiency. The findings are now published in *Nature Communications* and a patent application has been filed based on this work.

Background

Staphylococcus aureus Cas9 (SaCas9) is a great candidate for in vivo gene therapy due to its small size allowing packaging into adeno-associated viral vectors to be delivered into human cells for therapeutic applications. However, its gene-editing activity could be insufficient for some specific disease loci. Further optimizations of SaCas9 are crucial in precision medicine before it can be used as a reliable tool to treat human diseases. Such optimizations consist of boosting its efficiency and precision by altering the Cas9 protein. Standard protocol for modifying the protein entails saturation mutagenesis, where the number of possible modifications that could be introduced to the protein far exceeds the experimental screening capacity of even the state-of-art high-throughput platforms by orders of magnitudes.

In this work, the research team explored if combining machine learning with structure-guided mutagenesis library screening could enable the virtual screening of many more modifications to accurately identify the rare and better performing variants for further in-depth validations.

Research findings

The research team tested the machine learning framework on several previously published mutagenesis screens on Cas9 variants and illustrated that machine learning could robustly identify the best performing variants by using merely 5-20% of the experimentally determined data.

The Cas9 protein contains several parts, including protospacer adjacent motif (PAM)-interacting (PI) and Wedge (WED) domains to facilitate its interaction with the target DNA duplex. The research team coupled the machine learning and high-throughput screening platforms to design activity-enhanced SaCas9 protein by combining mutations in its PI and WED domains surrounding the DNA duplex bearing a (PAM). PAM is essential for Cas9 to edit the target DNA and the idea was to reduce the PAM constraint for wider genome targeting whilst securing the protein structure by reinforcing the interaction with the PAM-containing DNA duplex via the WED domain.

In the screen and subsequent validations, the researchers identified new variants, including one named KKH-SaCas9-plus, with enhanced activity by up to 33% at specific genomic loci. The subsequent protein modeling analysis revealed the new interactions created between the WED and PI domains at multiple locations within the PAM-containing DNA duplex, attributing to KKH-SaCas9-plus's enhanced efficiency.

Research significance

Structure-guided design has been dominating the field of Cas9 engineering; however, it only explores a small number of sites, amino-acid residues, and combinations. In this study, the research team showed that screening with larger scale and less experimental efforts, time and cost can be conducted using the machine learning-coupled multi-domain combinatorial mutagenesis screening approach, which led them to identify a new high-efficiency variant KKH-SaCas9-plus.

'This approach will greatly accelerate the optimization of Cas9 proteins, which could allow genome editing to be applied in treating genetic diseases in a more efficient way,' said Dr Alan Wong Siu-lun, Assistant Professor of the School of Biomedical Sciences, HKUMed.

Source:

The University of Hong Kong

Journal reference:

Thean, D.G.L., *et al.* (2022) Machine learning-coupled combinatorial mutagenesis enables resource-efficient engineering of CRISPR-Cas9 genome editor activities. *Nature Communications*. doi.org/10.1038/s41467-022-29874-5.
