Sheila Zhang
12th Grade
Sponsor: Julie Baker
Category: Medicine and Health

Abstract

Rice CRISPR Treat: A Method to Improve Detection of Off-Target CRISPR-Cas9 Activity Using Deep Sequencing Data and Homology Targeted Integration

CRISPR-Cas 9 is a revolutionary genome editing tool with an unmatched ability to silence and substitute desired genes across a wide range of cell types. However, it is prone to off-target edits, edits that occur when CRISPR binds to incorrect genome sites. Such edits often result in unintended changes to the genome and can produce harmful genetic mutations. Current research has applied wet-lab techniques to improve CRISPR's binding accuracy, but no methods have focused on the use of artificial intelligence or computational science. Thus, the use of a computationally powered framework to remotely identify sites of off-target editing and efficiently improve the editing accuracy of CRISPR technology would enable scientists to continually advance in genetic research.

The model was developed in a four-stage process. First, genome sequencing data was compiled from the National Center for Biotechnology Information and categorized into training and testing sets. Second, a next-generation sequencing analysis of the data was performed in order to recognize any abnormality seen across the genome. Third, data of the off-target genomic sequences were inputted into the model for testing. Fourth, using the off target-data, the model was trained to compare genomic abnormalities identified in the sequencing analysis to the off-target data.

The model was evaluated based on its ability to correctly identify instances of off-target editing using a system known as Precision and Recall, where the number of true/false positives and true/false negatives were recorded. Additionally, an ROC curve was plotted to visually display the degree to which the model could distinguish between the two classes. The finalized prototype yielded a significantly greater proportion of true positives/negatives than false positives/negatives, as well as a classification accuracy of approximately 84%, compared to CRISPR's accepted accuracy rate of approximately 60%.

With CRISPR technology in its infancy, the research encompassed in this project form a pathway to increased speed in genome screening, tools to prevent genetic diseases, and greater accuracy in gene therapy frameworks. This research further suggests that computational frameworks are a potent means of exciting advances in the gene editing field, widening the realm of CRISPR's capabilities.