CROTON: An Automated and Variant-Aware Deep Learning Framework for Predicting CRISPR/Cas9 Genome Editing Outcomes

Li, Victoria (School: Hunter College High School)

CRISPR/Cas9 is a revolutionary gene-editing technology that has been widely utilized in biology, biotechnology, and medicine. CRISPR/Cas9 editing outcomes depend on local DNA sequences at the target site and are thus predictable. However, existing prediction methods are dependent on both feature and model engineering, which restricts their performance to existing knowledge about CRISPR/Cas9 editing. Herein, deep multi-task convolutional neural networks (CNNs) and neural architecture search (NAS) were used to automate both feature and model engineering and create an end-to-end deep-learning framework CROTON (CRISPR Outcomes Through cONvolutional neural networks). The CROTON model architecture was tuned automatically with NAS on a synthetic large-scale construct-based dataset before it was tested on an independent primary T cell dataset. CROTON outperformed existing expert-designed models and non-NAS CNNs in predicting 1 base pair insertion and deletion probability as well as deletion and frameshift frequency. Interpretation of CROTON revealed local sequence determinants for diverse editing outcomes. CROTON was subsequently validated in laboratory experiments with hematopoietic stem and progenitor cells. CROTON was also utilized to assess how single nucleotide variants (SNVs) affect CRISPR/Cas9 gene-editing outcomes in two clinically relevant genes: the viral receptor CCR5 and immune checkpoint inhibitor PDCD1. To extend this analysis, a database describing the impact of SNVs on CRISPR/Cas9 editing outcomes across the human genome was created. This study is the first to systematically evaluate how genetic diversity affects CRISPR/Cas9 gene editing. CROTON and its associated database CROTONdb can be utilized for widely applicable clinical gRNA design.