Enhancing Gene Editing Safety: Leveraging 3D Genome Architecture and Epigenetics to Predict CRISPR-Cas9 Off-Target Mutations via Deep Learning

Jha, Shiven (School: Cumberland High School)

The clinical translation of CRISPR-Cas9 therapeutics is limited by the risk of unintended "off-target" mutations, which can cause errors like oncogenic chromosomal alterations. Current computational models designed to predict these off-target cleavage sites mainly rely on flat 1D DNA sequence alignment or basic local epigenetic tracks. These models fail to capture the 3D spatial geometry of the genome, ignoring the reality that dense chromatin loops can physically block the Cas9 endonuclease. To address this, I engineered TAG-Cas (Topological Accessibility Gating), a novel Triple-Branch deep learning architecture featuring a 3D Feature Pyramid Network (FPN). TAG-Cas integrates 23bp guide-target sequences, 200bp context with corresponding ATAC-seq scores, and multi-scale 3D topological geometry (11x11 Hi-C contact maps at 1kb, 10kb, and 50kb resolutions). These modalities are fused via a novel Multiplicative Biological Gating mechanism that mathematically enforces 3D accessibility. Trained and strictly cross-validated on the K562 crisprSQL database using gRNA-segregated splitting, TAG-Cas achieved a 0.916 PR-AUC. Ablation testing and McNemar's statistical analyses showed that multi-scale 3D structural accessibility is a major driver of off-target cleavage. Furthermore, Integrated Gradients mapped to the JASPAR database revealed that the model learned the regulatory "grammar" of the human genome, identifying master transcription factors (e.g., ZEB1, FIGLA) as structural Cas9 barriers. Ultimately, TAG-Cas advances the field of off-target prediction by showing that 3D topology is essential for accurate prediction, while offering a safer, more reliable computational framework to screen future CRISPR-based genetic therapies.

Awards Won:
Fourth Award of $600