AI-Optimized Multi-Gene Guide RNA Design for Salt-Tolerant Tomatoes

Jain, Dhruv (School: Lincoln Community School)

Current CRISPR guide design tools rank on single scores without multi-module integration. This project seeks to develop machine learning models to predict optimal CRISPR-Cas9 Guide RNA sequences for gene editing. Guide RNA candidates were scored across target genes. Four models were compared, all trained on 273 plant CRISPR guides with normalized CFD-overall labels. The results show that the gradient-boosted model achieved a five-fold cross-validated Spearman rho of 0.975 +/- 0.008, outperforming the equal-weight baseline (0.763 +/- 0.057) by 27.8%. Results were validated with non-overlapping confidence intervals; 69.1% of scored candidates received the worst off-target score. Thompson sampling over 1,000 rounds selected 25 final guides (5 per gene), with 18 of 23 unique candidates overlapping pure composite ranking, and 5 exploration picks promoted by Bayesian uncertainty. In conclusion, learned multi-module integration substantially improves guide ranking beyond single-score approaches, and the pipeline architecture generalizes to other crops and stress traits. No existing tool combines learned scoring integration with Bayesian selection.

Awards Won: