A computational approach for genome editing using CRISPR/Cas

Thaddeus D. Seher, Namkha Nguyen, Diana Ramos, Priyanka Bapat, Clarissa J. Nobile, Suzanne S. Sindi, and Aaron D. Hernday

Abstract

Typical genome editing using CRISPR/Cas-induced homology-directed repair has shortcomings:

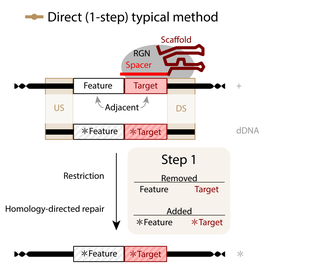
(1) Cannot edit a locus that lacks a target for RNA-guided nucleases, and cannot make homozygous edits if the locus has allelic variation
(2) Target quality is not comparable across experiments
(3) Difficult to design cost-effective PCR assay for verification

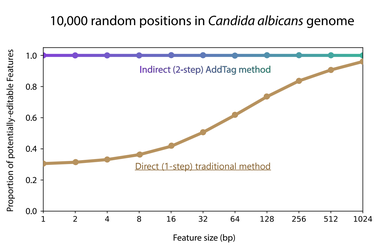
We present the AddTag method and AddTag software as solutions to these problems, and validate them in the *Candida albicans* biological system.

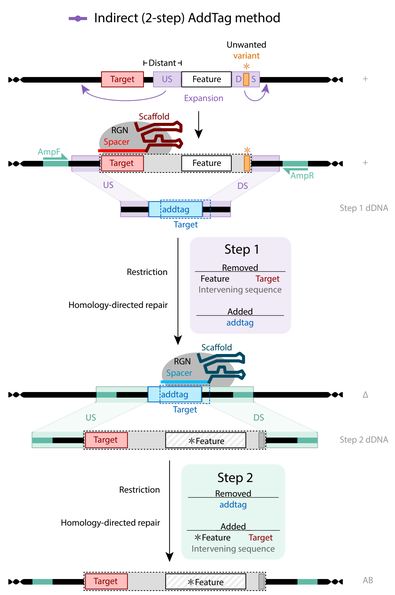
(1) Editing problematic loci

Direct editing requires the Target to be within- or adjacent to- the Feature being edited. Indirect editing allows for use of Targets far from the Feature and donor DNA (dDNA) homology arms without unwanted variants. Step 1 replaces the Feature and Target with an `addtag` sequence that encodes for an intermediary Target. Step 2 returns the deleted sequence with any desired modifications.

Use of the AddTag (2-step) method allows for edits to nearly the entire genome.







UNIVERSITY OF CALIFORNIA MERCED
SCHOOL OF NATURAL SCIENCES

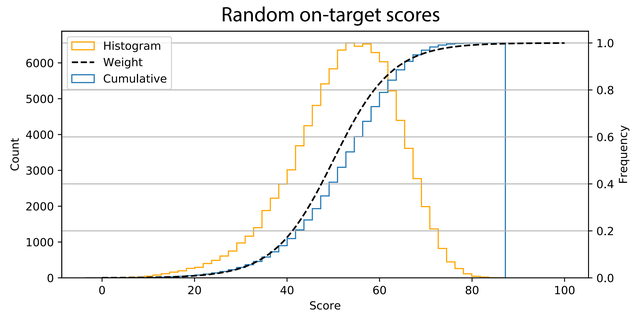
(2) Target quality across experiments

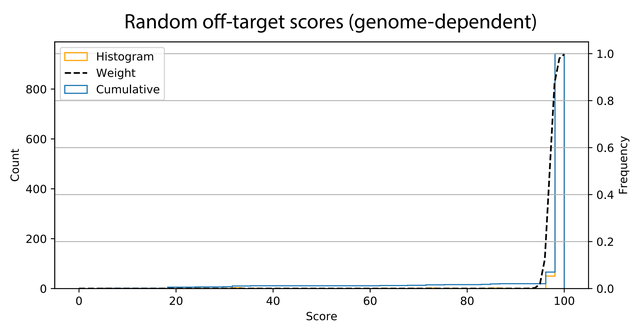
Equation for converting genome-dependent target score (*s*) to a genome-independent weight (*w*):

$$\theta = \{x, slope\}$$

$$w(s|\theta) = \frac{1}{1 + slope^{x-s}}$$

The sigmoid function allows for a low parameter (θ) approximation of the cumulative distribution function built from the scores of random nucleotide sequences (*seq*):





The importance of multiple scoring algorithms are balanced using their product:

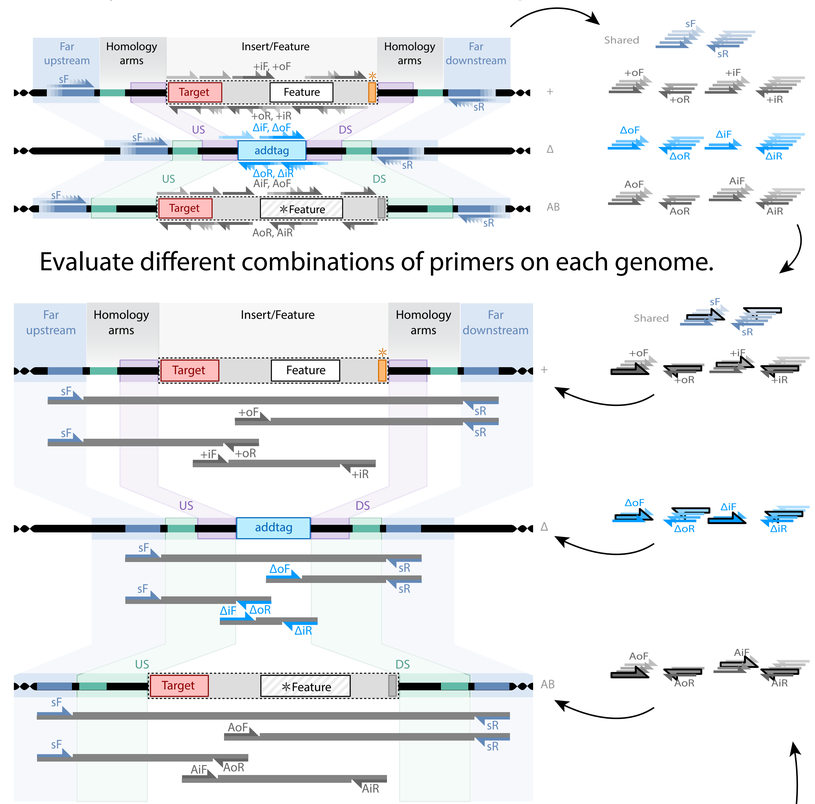
$$W(seq) = \prod_{a}^{A} w_a(s_a(seq)|\theta_a)$$

The final weight (*W*) assumes independance of each scoring algorithm ($a \in A$), and represents the relative utility of the sequence as a RNA-guided nuclease target.

(3) Designing PCR verification primers

Use simulated annealing (a type of genetic algorithm) to identify primer sequences with compatible thermodynamic properties.

Identify potential primers within specific regions



Finally, report predicted amplifications for each genome editing step

Availability

Manuscript under review
github.com/tdseher/addtag-project
@tdseher
tseher@ucmerced.edu