

S2 Text: Processing of the data used for validation.

Cho et al. [15]

Cho et al. used deep sequencing to assess the cleavage efficacies of 116 genomic sites that could potentially be cleaved by ten sgRNAs. These sites were selected such that they retain up to three mismatches from their corresponding sgRNA.

Wang et al. [22]

Wang et al. used deep sequencing to assess the cleavage efficacies of 54 genomic sites that could potentially be cleaved by two sgRNAs. These sites were selected such that there could be up to one mismatch at the first ten PAM-proximal nucleotides, and no more than four mismatches overall.

External data transformation

Similar to the data in our training dataset, these data represent indel formation frequencies. Therefore, we applied log transformation to them. To represent the data (S8-S9 Figs), we divided both datasets by their maximal values accordingly.