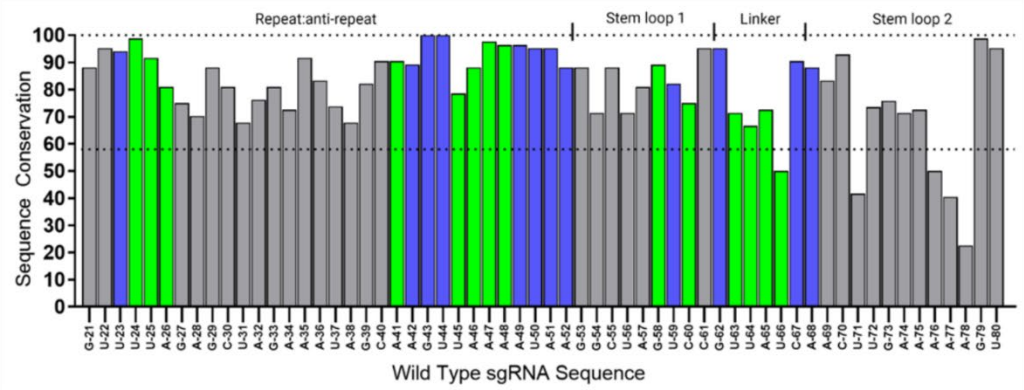
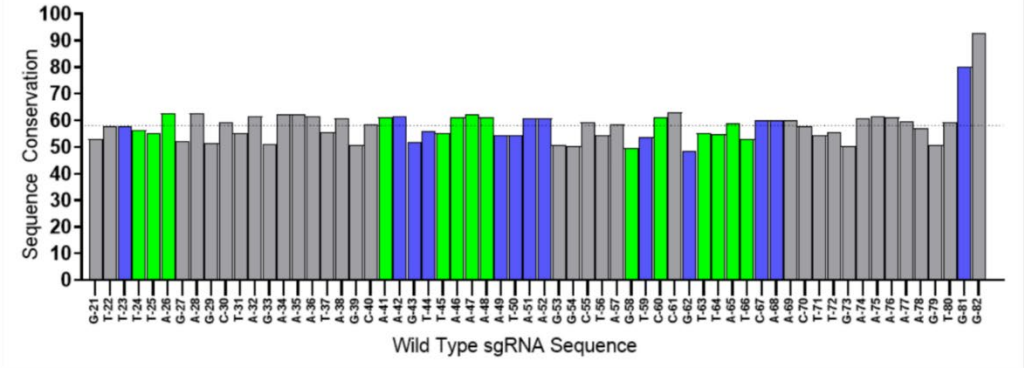
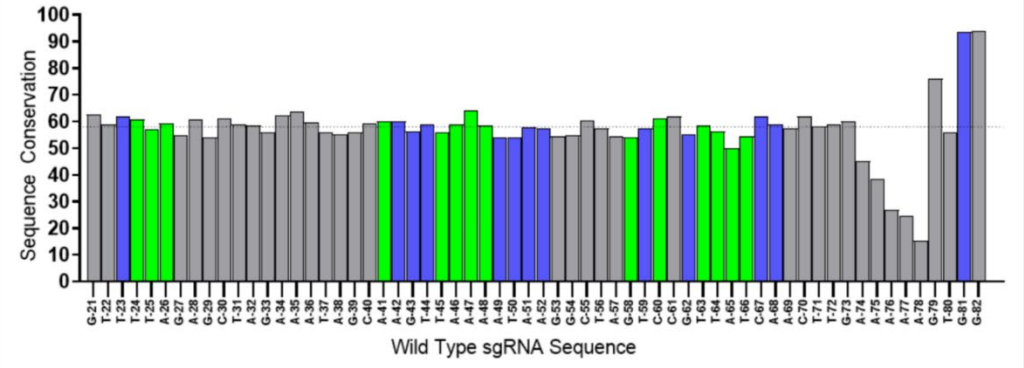
Functional Sequences



Round 1



Round 3



Round 5

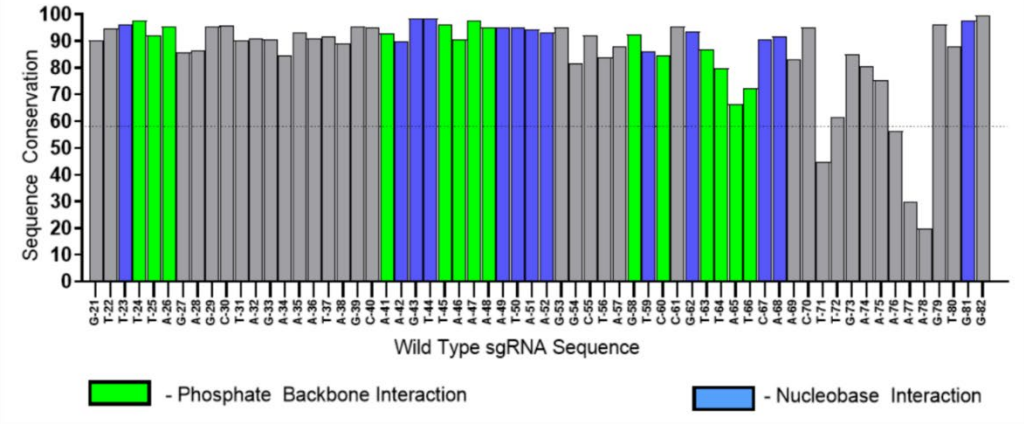


Figure S1. Selected sgRNA variation and correlation with Cas9 protein interactions, related to Figure 4. Ninety functional sgRNA variants were assessed for sequence identity using Geneious. The frequency of the WT base at each position of are plotted showing degree of conservation to the WT sgRNA sequence (at bottom, labeled with positions on the guide). Residues involved in interactions with the Cas9 protein are colored based on previous analysis [S1]. In green are positions which contact the protein through phosphate backbone interactions. In blue are residues involved in direct interaction between the nucleobase and the Cas9 protein. Nucleotide positions that do not participate in direct Cas9 interactions in the crystal structure are shown in gray. For reference, 100% conservation is marked across the top, and the 58% randomization of the beginning pool is shown as a dotted line. Deviations from the consensus at each position from the WT sgRNA sequence shows tolerance for alternate bases, especially in stem loop-2, where positions U71, A78, and A79 were replaced with G71, C78, and C79 at frequencies of 48%, 47%, and 43%, respectively.

For comparison, analogous sequence analyses of Rounds 1 (136,762 sequences), 3 (63,788 sequences), and 5 (8,009) are also presented, showing relatively even distribution of sequences in Rounds 1 and 3 and more defined sequences by Round 5.

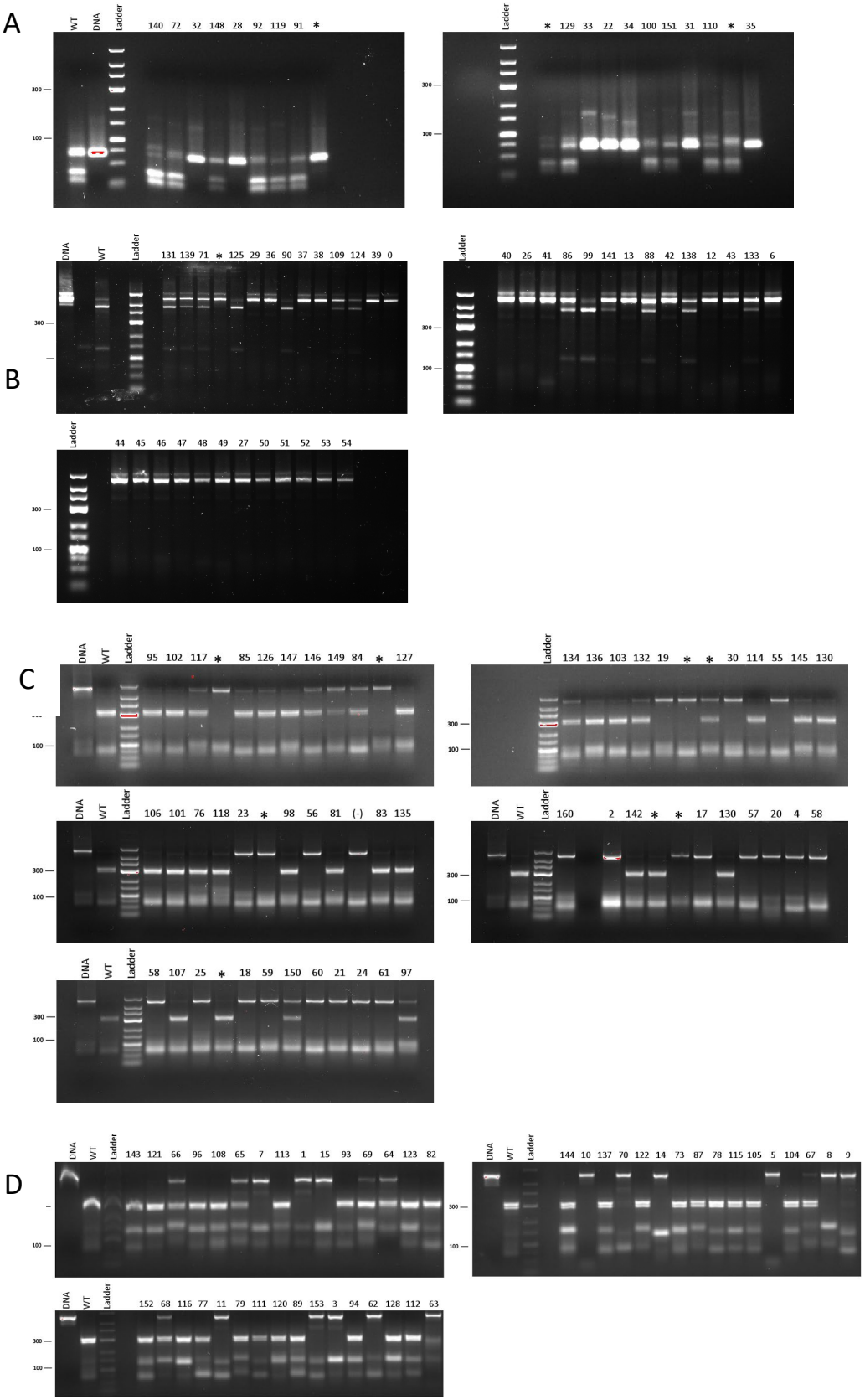


Figure S2. sgRNA variants from BLADE SELEX are largely capable of Cas9 mediated cleavage.

related to Figure 5. A.) Representative clones following the TdT functional screens were complexed with Cas9 and incubated with Substrate 1, as described in the Methods. The reactions were treated with proteinase K, run on 3% LE-agarose gels stained with SYBR SAFE, and imaged using BioRad Image Lab software. Cleavage of the 75 bp Substrate 1 is detected by the appearance of lower 30 and 35 bp bands. B-D.) Representative clones were complexed with Cas9 and incubated with Substrate 2. Cleavage of the 600 base pair Substrate 2 is detected by the appearance of a lower 300 bp band. Of the 153 clones analyzed from the functional selection, 89 demonstrated some degree of cleavage. The * indicates replicate samples, and the (-) indicates a negative control RNA. Gel images have been cropped to remove extra gel spaces.

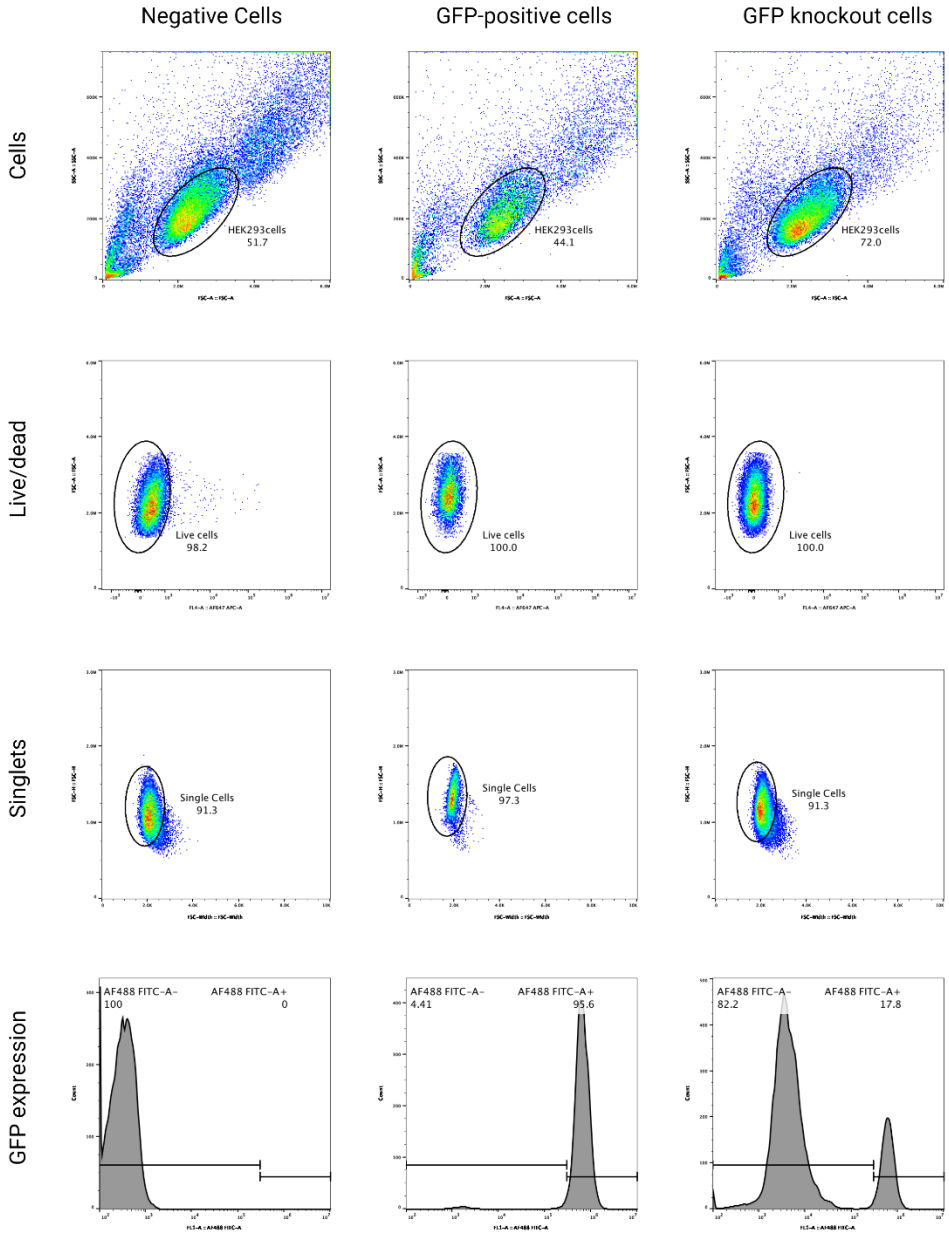


Figure S3. Example of gating strategy for flow cytometry experiments, related to Figure 5. Cell populations were gated away from debris and aggregates (top). This subpopulation was further gated to remove dead cells based on SYTOX Red staining (second row). Live cells were further gated to remove aggregates (third row). A double gate was used to distinguish between cells that were GFP-positive and those with knocked out GFP expression (bottom). Shown are examples of non-GFP expression parental HEK293 cells (left), GFP-expressing HEK293 cells (middle) and a sample of HEK293-GFP cells with knockout of GFP (right). We note that for knockout populations, GFP expression usually did not go

down to non-GFP expressing levels. For each experiment, gates were kept fixed across the samples analyzed.

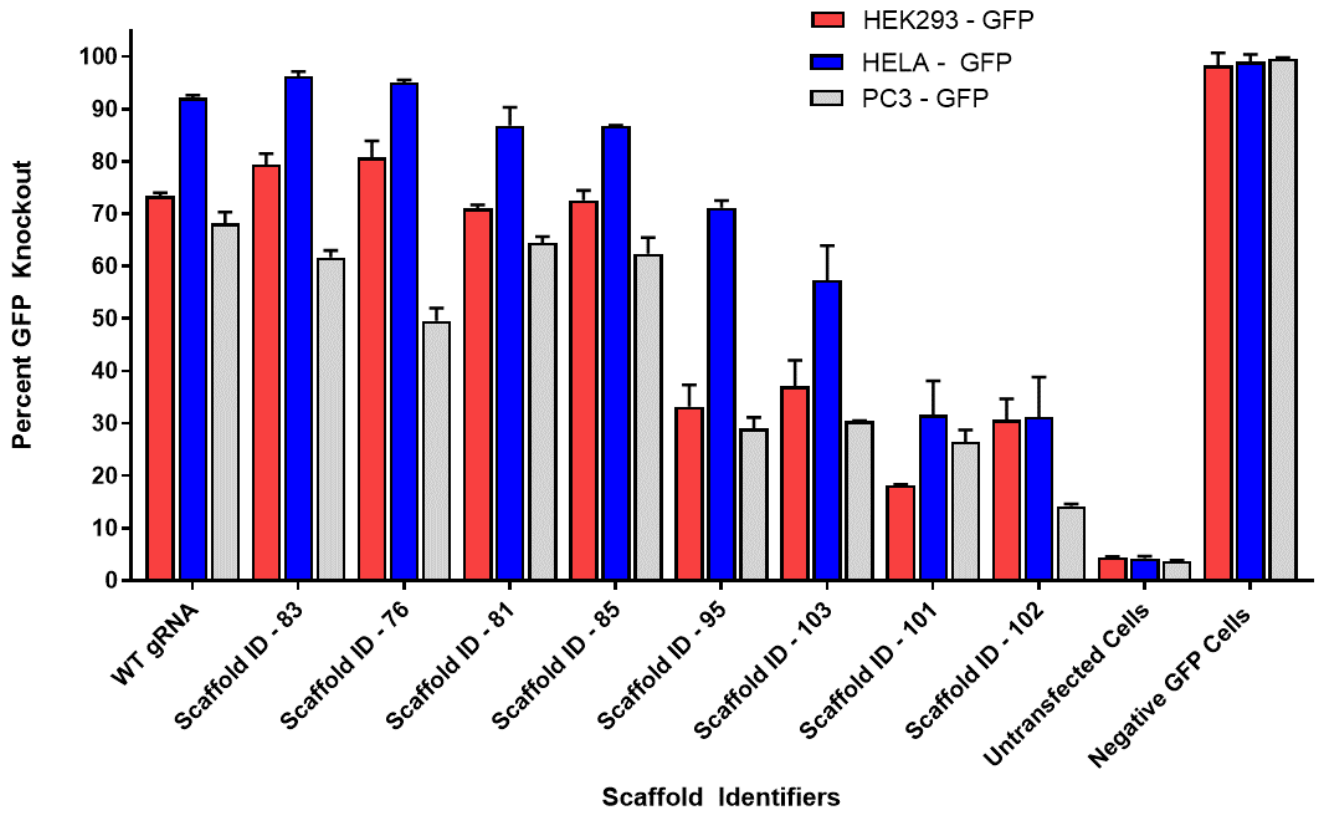
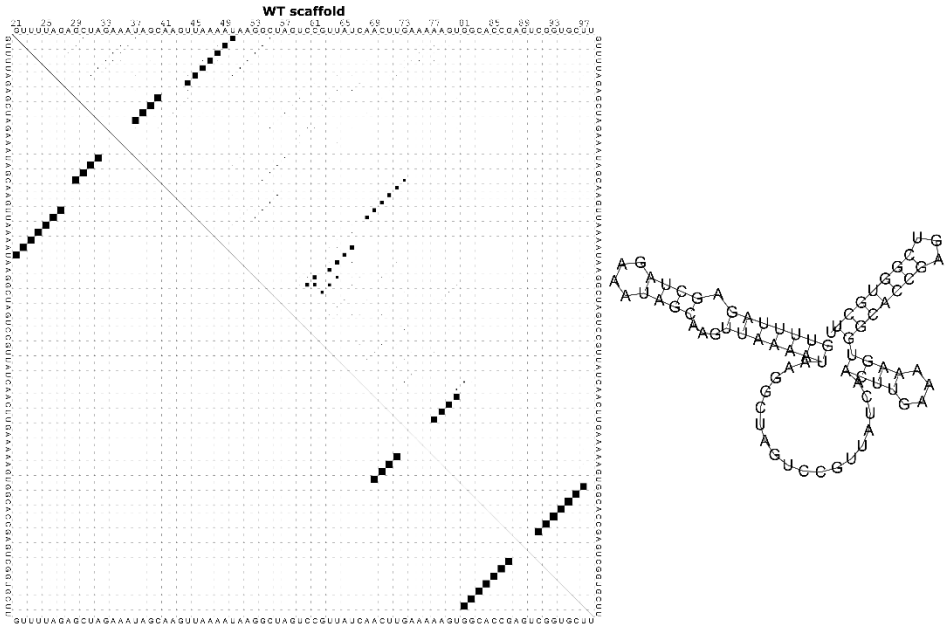
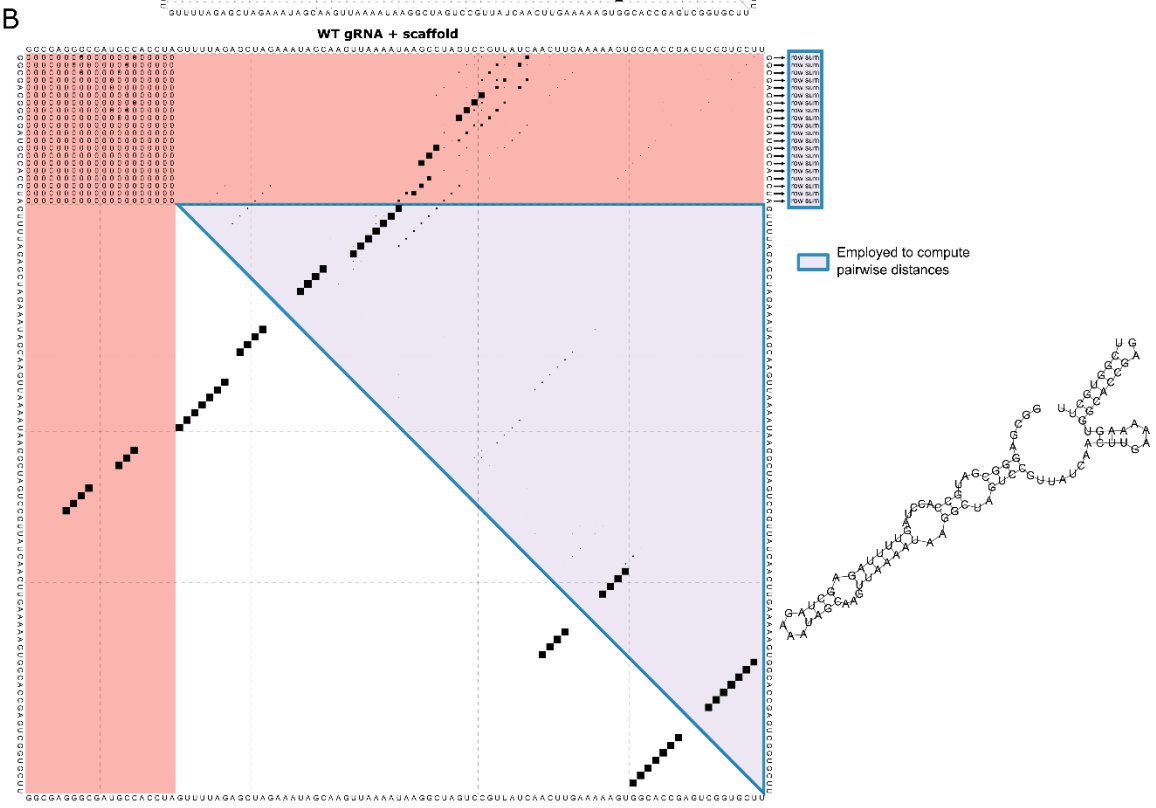


Figure S4. Selected sgRNA variants show different levels of effectiveness in knocking out GFP in cells, related to Figure 5. Selected sgRNA variants were tested for knockout of GFP in 3 cell lines constitutively expressing GFP. Of note, the HeLa-GFP cells express a destabilized, short-lived version of GFP. GFP expression was determined by flow cytometry on Day 6 post-transfection, and data are shown as the percentage of cells that were GFP negative in the population. GFP knockout was also assessed for untransfected cells, a transfection control (only Cas9, no sgRNA) and the parental non-GFP expressing cells (Negative). Data represent greater than 2 replicates of each sample.

A


WT scaffold

B


WT gRNA + scaffold
Employed to compute pairwise distances

C

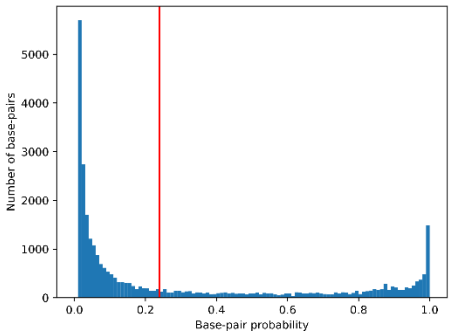

Number of base-pairs
Base-pair probability
0
1000
2000
3000
4000
5000
0.0
0.2
0.4
0.6
0.8
1.0

Figure S5. RNAfold base pairing probability matrix and secondary structure of the WT scaffold without (A) or with (B) the gRNA., and base pairing probability threshold, related to Figures 5 and 6. A-B.) In each matrix, the upper triangle displays a dot plot of the base pairing probabilities while the bottom triangle shows the minimum free energy structure. The gRNA + scaffold matrix is annotated as follows to describe which elements of the matrix are involved in the computation of distances between matrices: rows and columns pertaining to the gRNA are highlighted in red; cells showing the spacer's self-folding are filled with zeros; the sum-up of the probabilities in the top 20 rows is shown at the right of the plot; and all the elements used to compute the distances are colored in blue. C.) Histogram showing the number of cells in all of the base pairing probability matrices that fall in a certain range of base pairing probability values. To simplify the visualization, 1,572,522 base pairing probability values below 0.01 are excluded from the plot. A red bar indicates the threshold of base pairing probabilities used in the computation of pairwise distances (0.24).

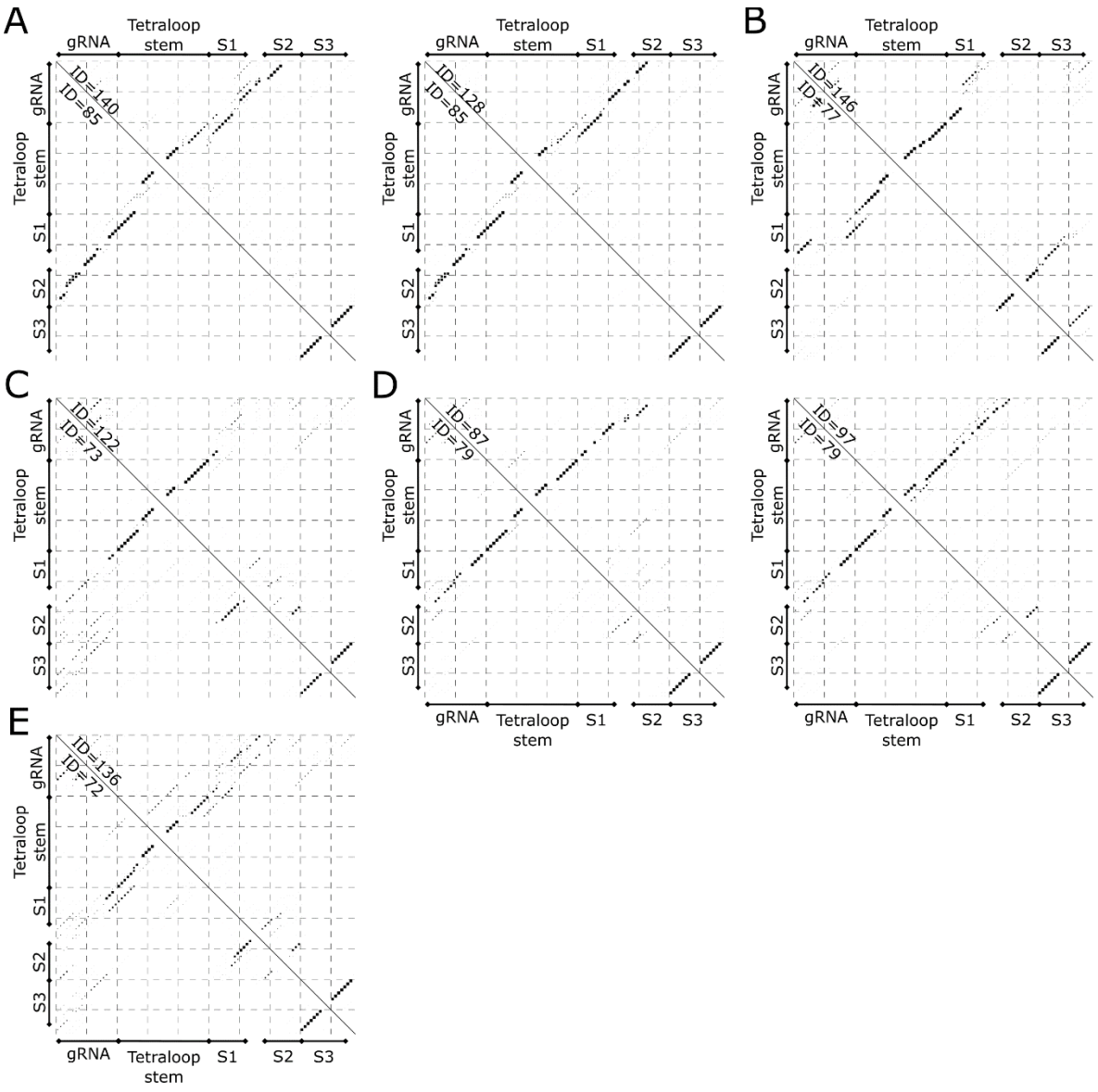


Figure S6. Comparison between base pairing probability matrices of highly efficient sgRNAs with alternative functional structures and closest neighbors in the hierarchical clustering, related to Figures 5 and 6. Base pairing probability matrices of sgRNAs (upper triangle) neighboring highly efficient sgRNAs (bottom triangle) with alternative functional structures in the dendrogram of the hierarchical clustering of sgRNA structures, displayed as dotplots. The sgRNA IDs, corresponding to the IDs in Figure 5, are given along the diagonal. The Tetraloop Stem, Stem 1 (S1), Stem 2 (S2), and Stem 3 (S3) are annotated. Shown are the structural neighbors of A.) sgRNA ID = 85, B.) sgRNA ID = 77, C.) sgRNA ID = 73, D.) sgRNA ID = 879, and E.) sgRNA ID = 72.

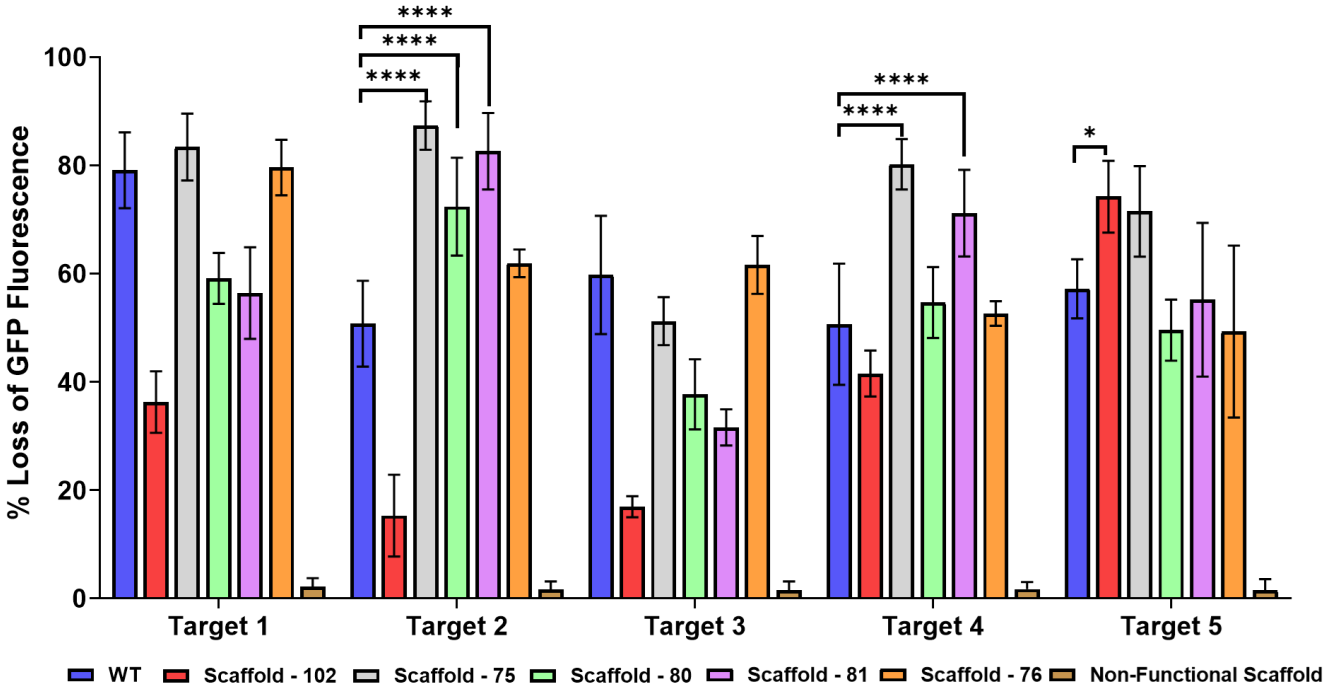


Figure S7. Selected variants demonstrate altered efficiencies when targeted to other sites, related to Figure 7. The wild type (WT) and 5 selected sgRNAs as well as a nonfunctional scaffold were retargeted to 4 other target sites on the GFP gene, and GFP knockout was assessed in HEK293-GFP cells. *GFP* Target 1 was the DNA site utilized during selection during BLADE SELEX. The degree of GFP knockout for each scaffold variant is shown relative to knockout due to the WT scaffold. Scaffolds are colored as indicated. One way ANOVA was used to compare each of the scaffolds for a given target to the WT of the same target. Scaffolds that display significant enhancements in editing efficiencies are denoted above. Stars represent p values of significance (* 0.0425; **** <0.0001).

Round #	Scheme	Target	Cas9 (µM)	RNA pool (nM)	DNA (µM)	Buffer
1	Nitrocellulose	Cas9	0.1	1	n/a	1*
2	Nitrocellulose	Cas9	0.1	1	n/a	1
3	Bead-based capture	Cas9 + DNA	0.1	1	0.001	2**
4	TdT + Bead=based capture	Cas9 + DNA	0.1	0.1	0.001	2
5	TdT + Bead=based capture	Cas9 + DNA	0.1	0.1	0.001	2

*Binding Buffer 1: 20 mM HEPES pH 7.4, 100 mM NaCl, 1 mM MgCl2, 0.001% BSA

**Binding Buffer 2: 50 mM Tris-Cl pH 7.9, 100 mM NaCl, 10 mM MgCl2, 0.001% BSA

Table S1: **Conditions for BLADE SELEX, related to Figure 3.** The sgRNA SELEX scheme winnowed variant diversity through standard nitrocellulose filter-based SELEX for RNP formation and binding to SpCas9 (Rounds 1-2), a magnetic bead-based selection for binding of the RNPs to target DNA (Round 3), and then a functional TdT-based selection for cleavage permissive sgRNA variants (Rounds 4-5). Concentrations of components and buffer conditions are as indicated.

SUPPLEMENTAL REFERENCES:

S1. Nishimasu, H., et al. (2014). "Crystal structure of Cas9 in complex with guide RNA and target DNA." Cell **156**(5): 935-949.