Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/29/eabb4054/DC1

Supplementary Materials for

A positive, growth-based PAM screen identifies noncanonical motifs recognized by the *S. pyogenes* Cas9

D. Collias, R. T. Leenay, R. A. Slotkowski, Z. Zuo, S. P. Collins, B. A. McGirr, J. Liu, C. L. Beisel*

*Corresponding author. Email: chase.beisel@helmholtz-hiri.de

Published 15 July 2020, *Sci. Adv.* **6**, eabb4054 (2020)
DOI: 10.1126/sciadv.abb4054

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/29/eabb4054/DC1)

Supplementary Table

Table S1. Binding energies from simulations, summary of targeting activity, list of strains, plasmids, and oligos, and p-values separated in different tabs.

Tab 1: Binding energies of R1333 and R1335 with different PAMs in the bulge-free DNA systems. The binding energies were calculated by the MM-GBSA approach. For a direct comparison, only the recognized dinucleotide within the PAM has been included into calculations. Values represent the mean and standard deviation calculated over a structural ensemble of 2,000 snapshots.

Tab 2: Binding energies of R1333 and R1335 with the canonical NGG PAM in the bulged DNA systems. The binding energies were calculated by the MM-GBSA approach. For a direct comparison, only the recognized dinucleotide within the PAM has been included into calculations. Values represent the mean and standard deviation calculated over a structural ensemble of 10,000 snapshots.

Tab 3. Summary of targeting activity using NYGG PAMs. Bolded entries indicate when the 20th nt of the target and the 1st nt of the PAM match.

Tab 4. List of strains, plasmids, and oligonucleotides used in this work.

Tab 5. List of p-values for two-tailed t-tests.

Supplementary Files

File S1. Interactive Krona plot for the PAM wheel in Figure 1C.

File S2. Interactive Krona plot for the PAM wheel in Figure S3.

Supplementary Figures

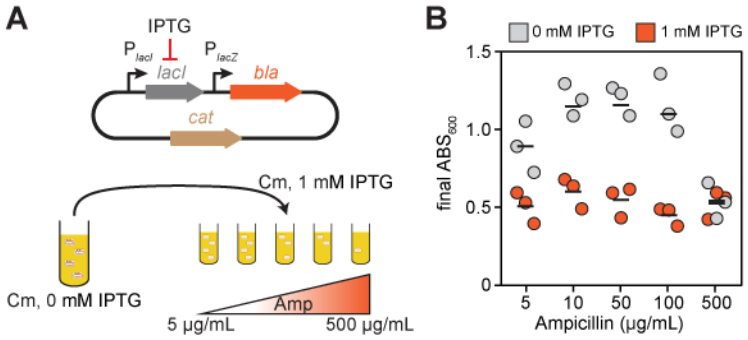

Fig. S1. Assessing a growth-based circuit conferring ampicillin resistance in the presence of a PAM. **A)** The *gfp* gene from the PAM-SCANR circuit was replaced with the *bla* resistance gene. *E. coli* cells harboring the circuit were challenged with a gradient of ampicillin selection in the presence or absence of the inducer IPTG. **B)** Growth in liquid culture following selection on different concentrations of ampicillin. Values represent final ABS_{600} measurements after overnight growth from independent experiments starting from separate *E. coli* colonies. Bars represent the mean of each set of triplicate measurements. The cells exhibited substantial growth even in the absence of IPTG and the presence of high ampicillin concentrations, leading us to discard this circuit as the basis of the growth-based PAM screen.

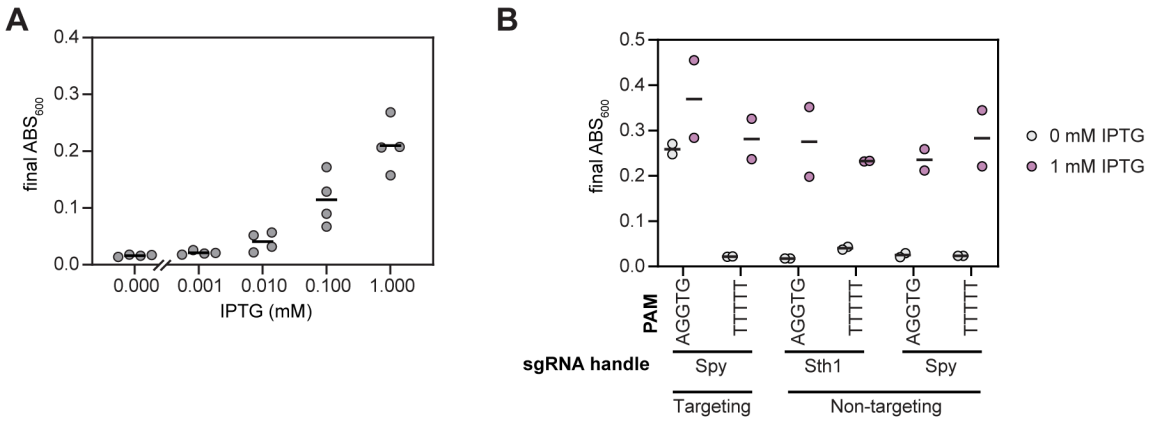

Fig. S2. Evaluating the properties and output of the growth-based circuit conferring catabolism of D-xylose. A) Cells with a non-targeting sgRNA were grown in D-xylose medium with the indicated concentration of IPTG. Bars represent the mean of each set of quadruplicate measurements. **B**) Cells were grown in the xylose medium with or without IPTG induction. Values represent final ABS_{600} measurements after overnight growth from independent experiments starting from separate *E. coli* colonies. Bars represent the mean of each set of duplicate measurements.

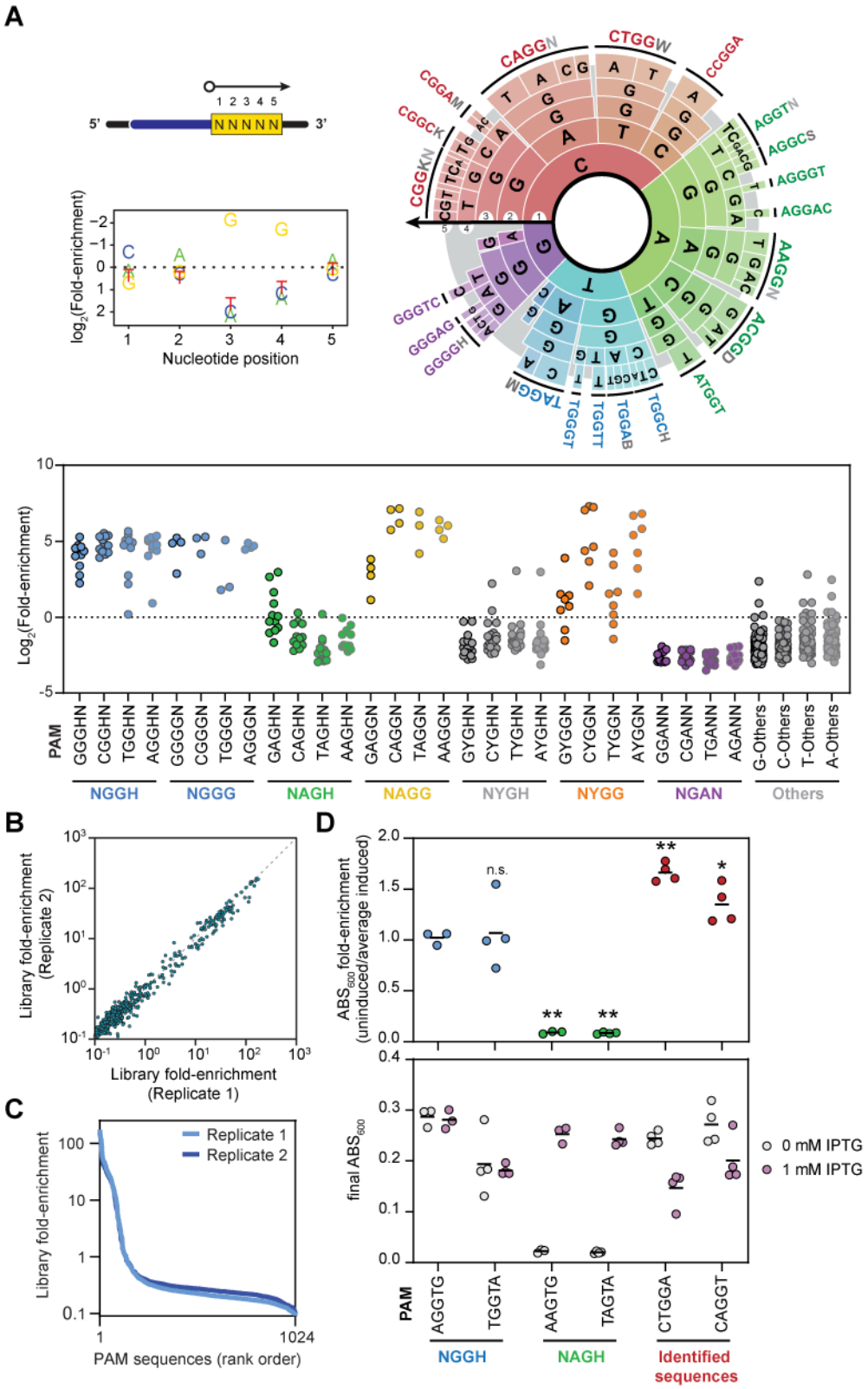

Fig. S3. See next page for the figure legend.

Fig. S3. Biological replicate of the growth-based PAM screen and validation of individual sequences in *E. coli*. A) Output of the biological replicate of the screen performed with Cas9. See Fig. 1 for more details. Each dot represents a distinct five-nt sequence within the library. See File S2 for an interactive version of the PAM wheel based on the Krona plot. **B)** Enrichment scores between biological replicates. Each dot represents a distinct five-nt sequence within the library. **C)** Library fold-enrichment for the biological replicates according to the rank-order of the PAM sequences. **D**) Assessing individual PAM sequences in *E. coli* using the growth-based genetic circuit with the optimized medium conditions. Top: ratio of final turbidity measurement. Bottom: individual turbidity measurements for the induced and uninduced cultures. The displayed PAM sequences were cloned in place of the original PAM library, and each was tested with a targeting sgRNA with and without IPTG induction. The AGGTG and AAGTG data represent the same data presented for those sequences in Fig. 2A. Statistical significance was calculated in comparison to the AGGTG sequence for using a two-tailed t-test assuming equal variance with a requirement value of 0.05 (*) or 0.01 (**). Bars represent the mean of at least three replicates starting from independent colonies.

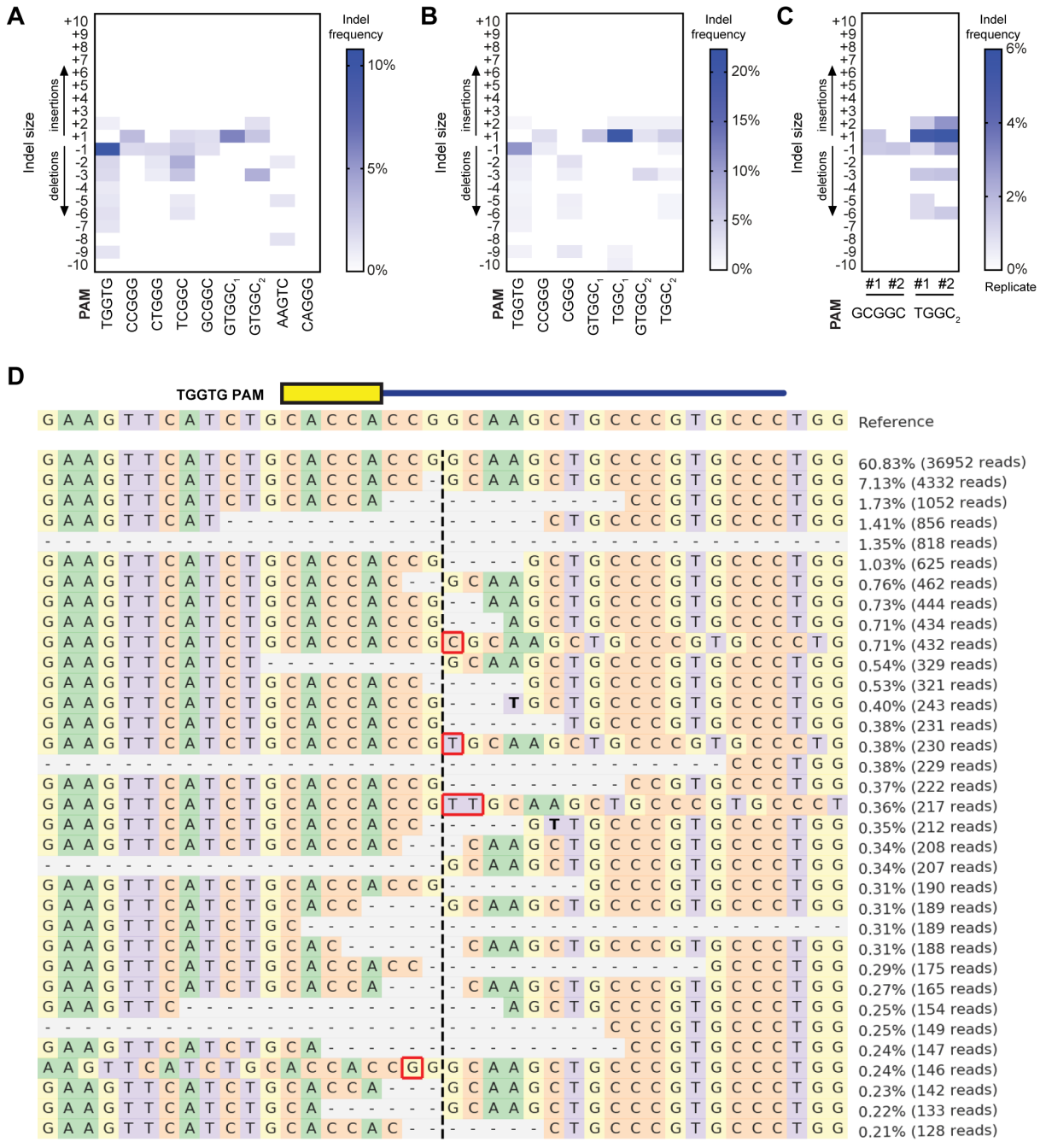

Fig. S4. See legend at end of figure panels.

Fig. S4. See legend at end of figure panels.

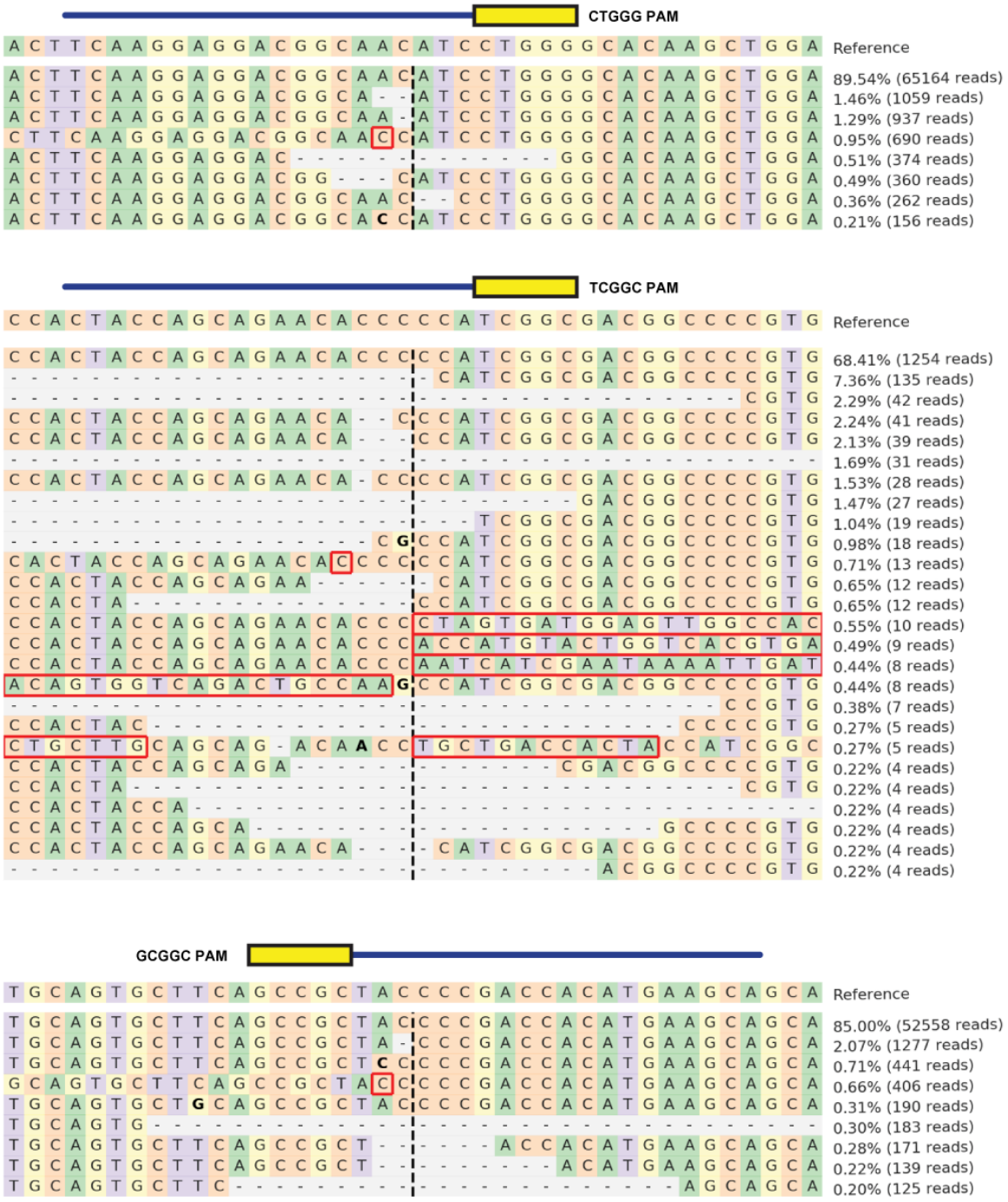

Fig. S4. See legend at end of figure panels.

Fig. S4. See legend at end of figure panels.

Fig. S4. See legend at end of figure panels.

Fig. S4. Indel formation upon targeting *EGFP* in U2OS cells with SpCas9. **A)** Indel size and frequency determined by TIDE for each target flanked by the indicated PAM sequence. **B)** Indel size and frequency determined by TIDE for targets shifted by 1 nt. The adjacent PAM sequence is shown, where the target was shifted by 1 nt to yield a YGG or NYGG PAM. **C)** Reproducibility of the indel distribution for two targets. Each sample represents an independent experiment starting with DNA transfection. **D)** Next-generation sequencing results analyzed by CRISPResso for each PAM motif in distinct locations.

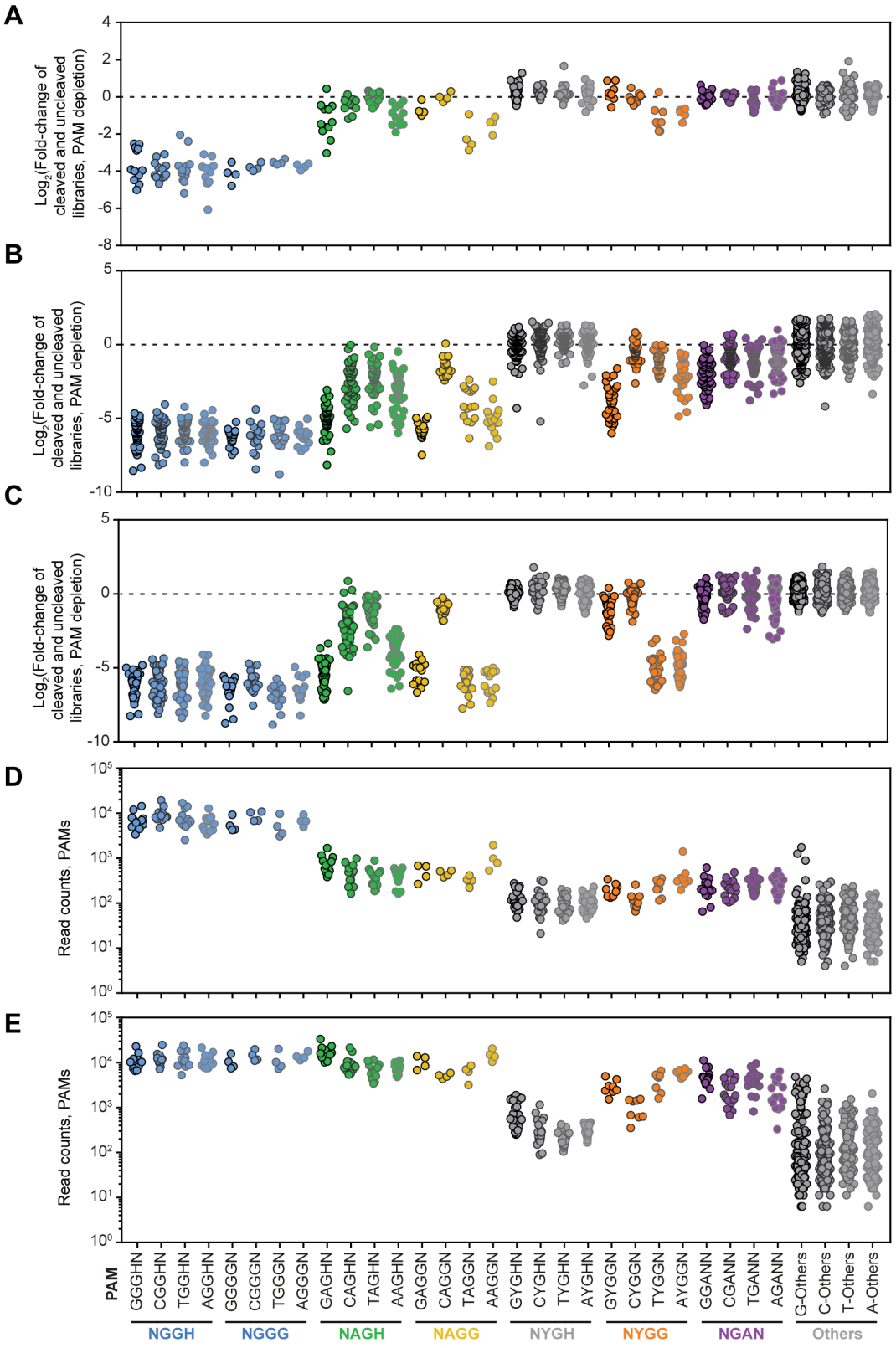

Fig. S5. See next page for the figure legend.

Fig. S5. Analysis of previously published, high-throughput PAM assays for Cas9 with at least 5N PAM libraries. A - C) Analysis of previous high-throughput PAM assays based on plasmid depletion in *E. coli (5, 12)*. A is based on one library in one study *(5)* while B and C are based on two libraries (Library 1 and Library 2, respectively) from another study *(12)*. Depleted sequences would be associated with more preferred PAM sequences. PAM assays were conducted with a 5N PAM library for A and a 6N PAM library for B and C. **D - E)** Analysis of previous high-throughput *in vitro* PAM assay with a 5N library and different concentrations of Cas9:crRNA:tracrRNA ribonucleoprotein complex *(15)*. D and E are based on the screen conducted with 0.5 nM and 50 nM Cas9, respectively. The resulting PAM sequences were grouped by the indicated motif.

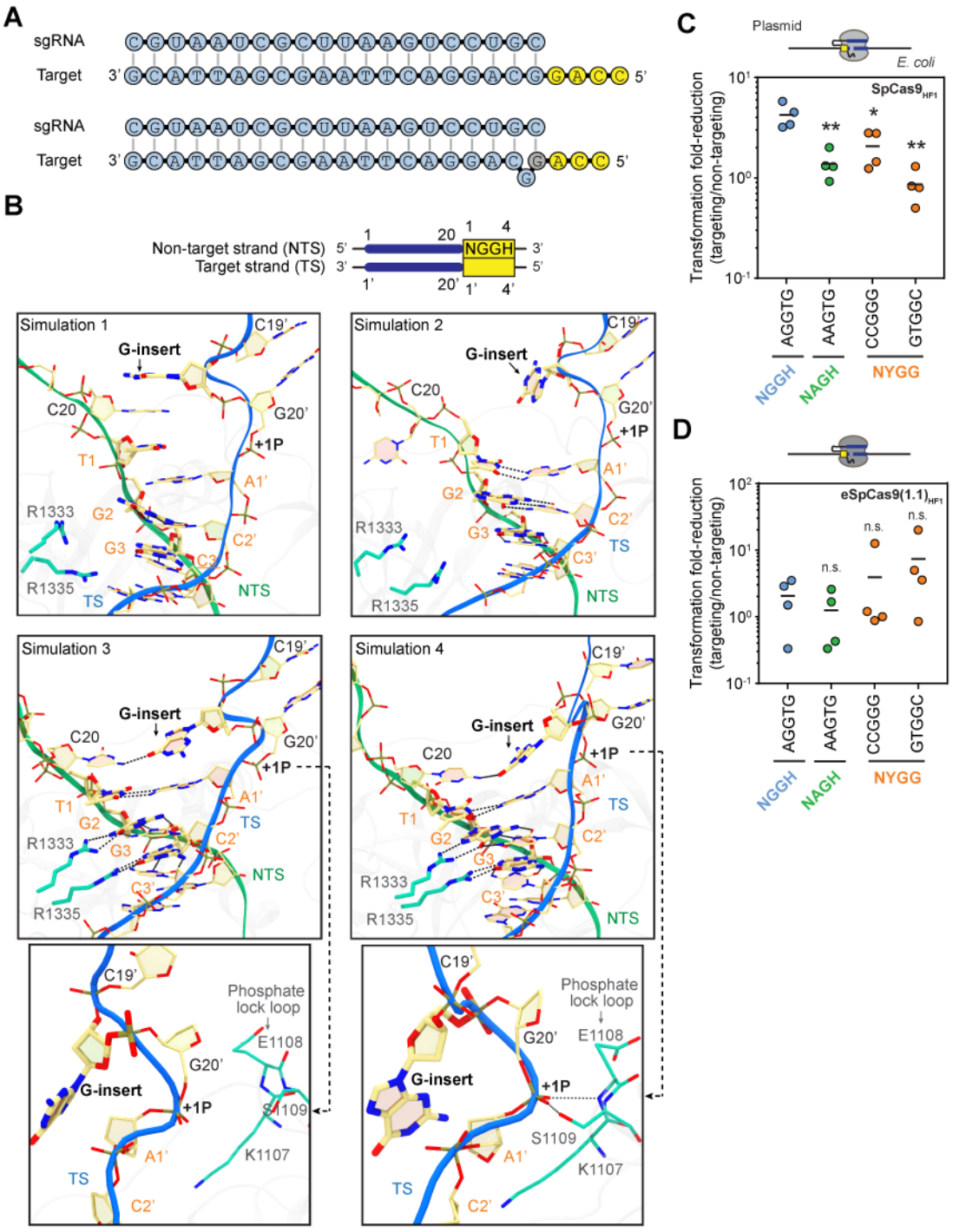

Fig. S6. Molecular dynamics simulations of Cas9 recognizing NYGG as a bulged target, and plasmid clearance by high-fidelity variants of Cas9. A) Comparing recognition of an NYGG PAM (GACC on the target strand) versus recognition of an NGG PAM (ACC on the target strand) with a one-nt bulge between the 19th and 20th nt of the target strand. For the bulge to occur, the 20th nt of the guide can base pair with the first nt of the NYGG PAM on the target strand. Only the R-loop formed between the guide and the

target strand are shown, and the PAM is highlighted in yellow. **B)** DNA bulge-induced structural perturbation observed across four simulation replicas. The extra G-nucleotide that introduces a bulge in DNA is denoted with an arrow. "**+1P**" represents the phosphodiester group immediately adjacent to the PAM duplex in the target strand. Below the bottom panels is a zoomed-in view of the phosphate lock loop-mediated interaction with **+1P**, where the nucleobases around **+1P** in the target strand are not displayed for clarity. The hydrogen bonds are denoted with dashed lines. In Simulations 1 and 2, the bi-arginine completely lost hydrogen-bond contacts with the PAM. In Simulation 3, **+1P** was disengaged from the phosphate lock loop that has an important role for dsDNA unwinding, although the bi-arginine maintained PAM recognition. In Simulation 4, both the PAM recognition and **+1P** interaction with the phosphate lock loop remained unchanged, although the bulged G-nucleotide is aligned in an orientation opposite to that in the other simulations, leading to a highly strained turn in the target strand. **C - D)** Poor plasmid clearance by $SpCas9_{HF1}$ and $eSpCas9(1.1)_{HF1}$ variants in *E. coli* in the absence of a non-selective outgrowth. See Fig. 2D for details. Values represent measurements from independent experiments starting from separate *E. coli* colonies. Bars represent the mean of each set of quadruplicate measurements. A value of 10^0 represents an equal number of transformants between the targeting and non-targeting sgRNAs. Statistical significance was calculated in comparison to the AGGTG sequence for each variant using a two-tailed t-test assuming equal variance with a requirement value of 0.05 (*) or 0.01 (**).

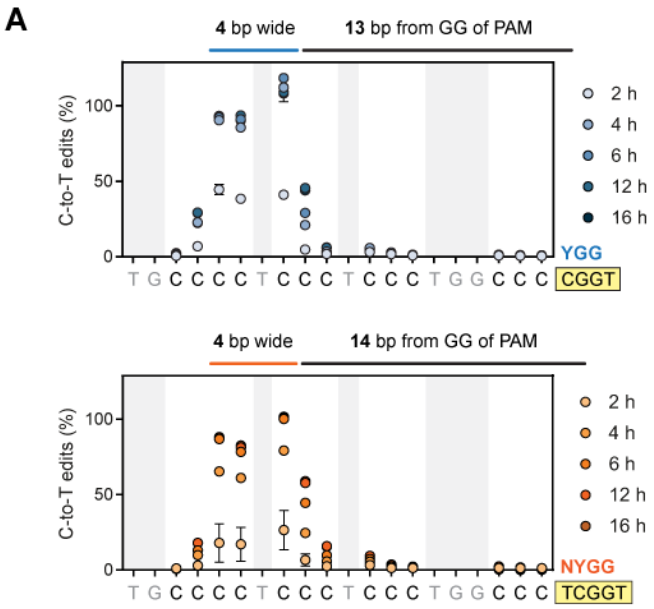

Fig. S7. See legend at end of figure panels.

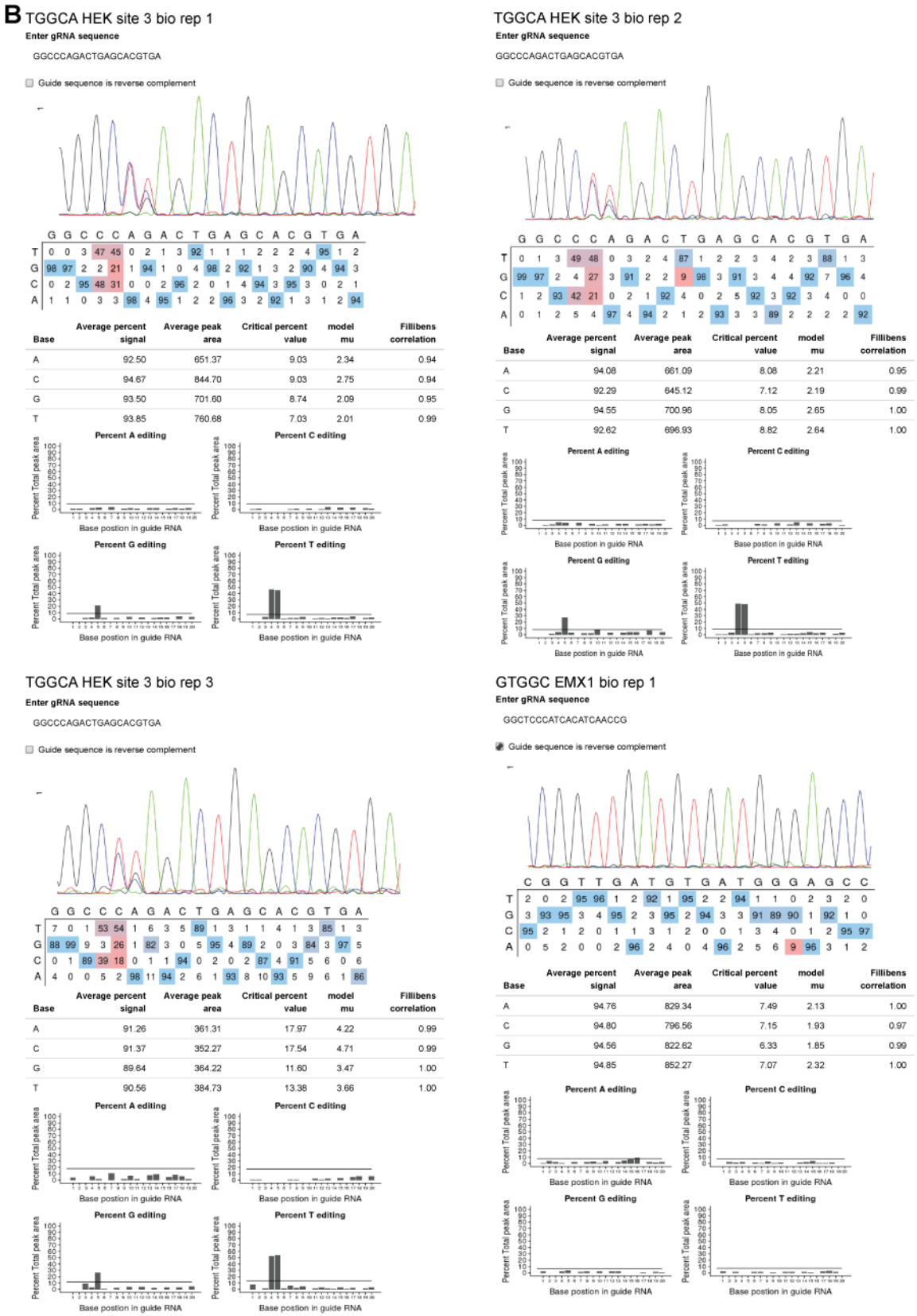

Fig. S7. See legend at end of figure panels.

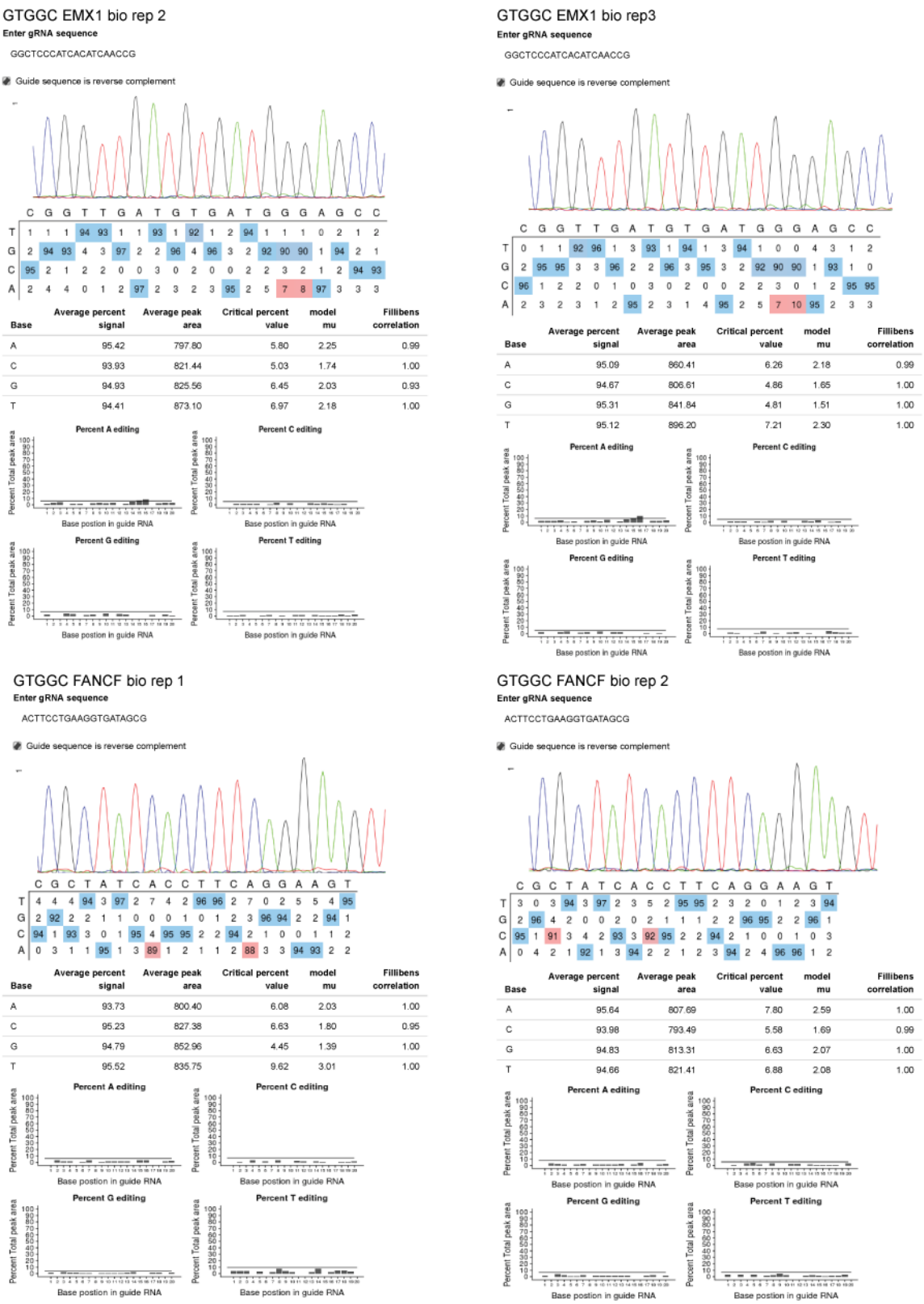

Fig. S7. See legend at end of figure panels.

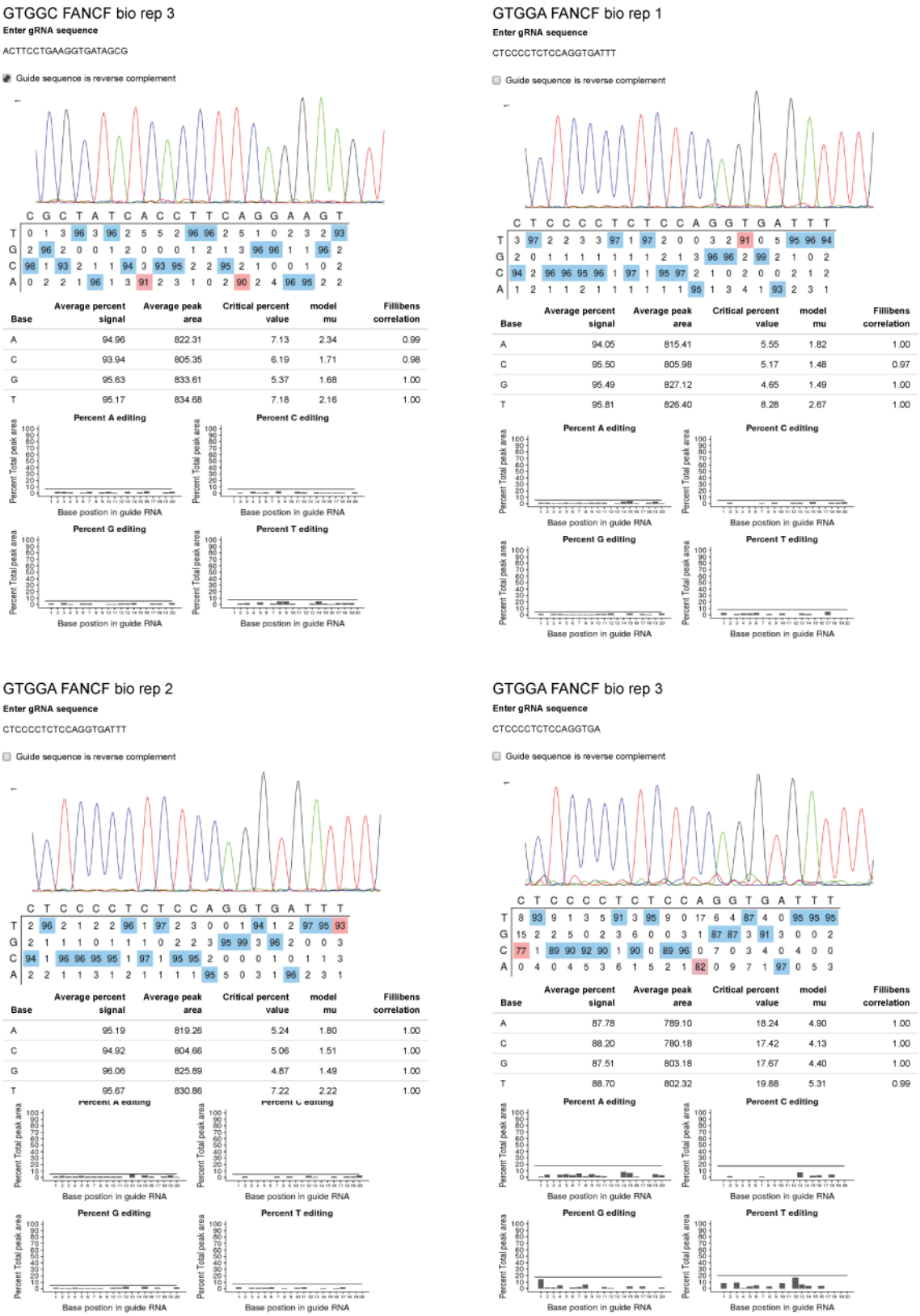

Fig. S7. See legend at end of figure panels.

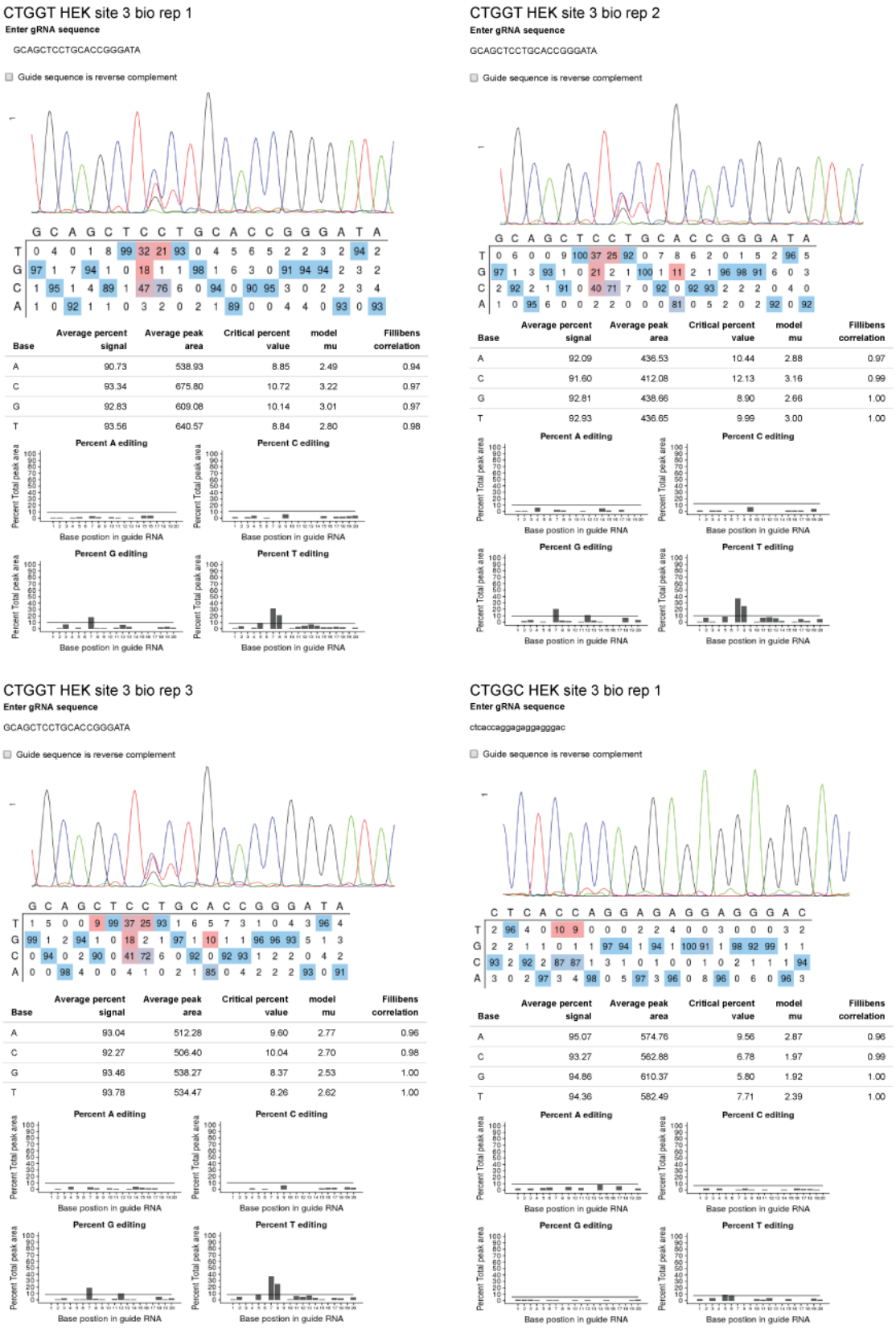

Fig. S7. See legend at end of figure panels.

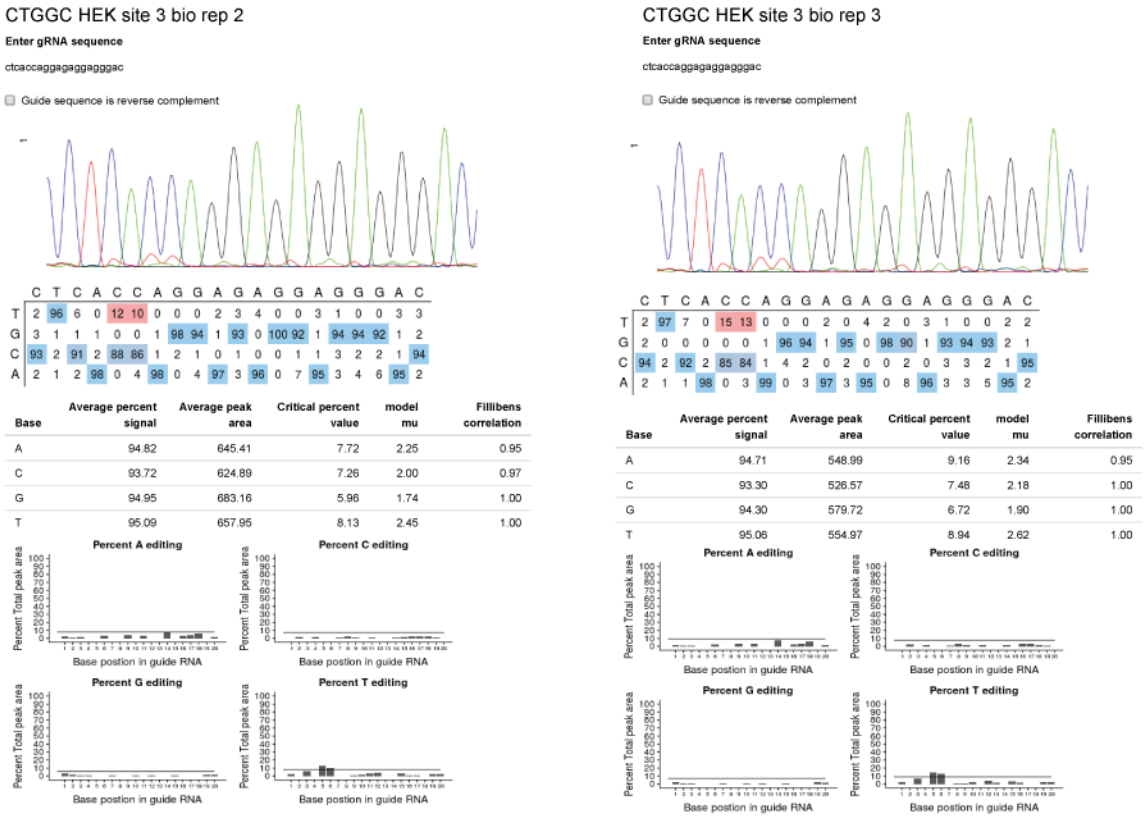

Fig. S7. Evaluating base editing with different PAMs in *E. coli* and at endogenous sites in human cells

A) Determining the base-editing locations in plasmid targets flanked by CGGT or TCGGT in *E. coli*. See Fig. 6C for more details. Values are in comparison to a non-targeting control and represent the mean and S.D. from three independent experiments starting from separate *E. coli* colonies. **B)** Base editing at different endogenous sites in HEK293T cells. The displayed output from EditR *(35)* was used to report the frequency of C-to-T edits. Analyses were conducted from three independent biological replicates starting from HEK293T cells seeded into independent wells.