

## **Supplementary Information**

### **A Cas9 with PAM Recognition for Adenine Dinucleotides**

Pranam Chatterjee<sup>1,2,\*</sup>, Jooyoung Lee<sup>3</sup>, Lisa Nip<sup>1,2</sup>, Sabrina R.T. Koseki<sup>1,2</sup>, Emma Tysinger<sup>1,2</sup>, Erik J. Sontheimer<sup>3</sup>, Joseph Jacobson<sup>1,2</sup>, and Noah Jakimo<sup>1,2</sup>

<sup>1</sup>MIT Center for Bits and Atoms, Massachusetts Institute of Technology, Cambridge, MA, USA

<sup>2</sup>MIT Media Lab, Massachusetts Institute of Technology, Cambridge, MA, USA

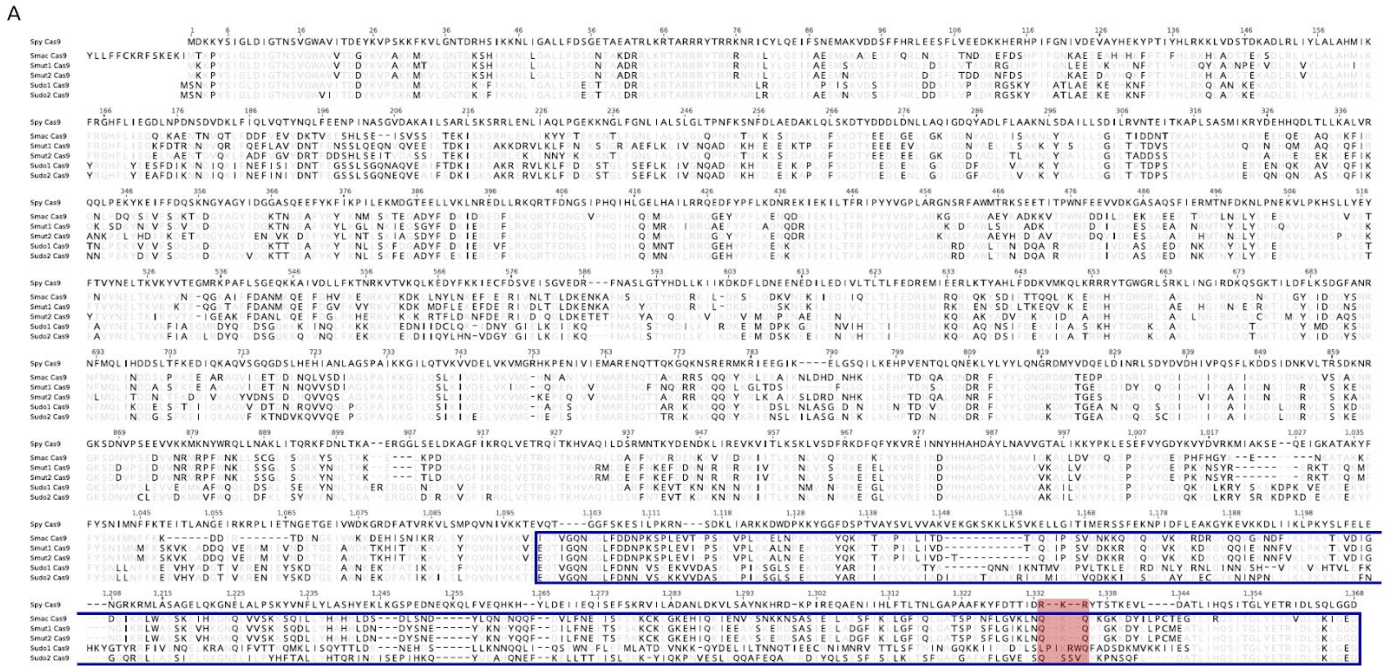
<sup>3</sup>RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA, USA

\*Corresponding author

## **Supplementary Figures**

1. Adenine nucleotide occurrences in human genome.
2. Comparisons of candidate orthologs.
3. CRISPR cassettes of candidate orthologs.
4. Flow cytometry gating strategy for PAM-SCANR assay.
5. tracrRNA determination for SpyMac.
6. Comparison of SpyMac to Cas12a as A-rich targeting platforms.





S = *Streptococcus*, py = *pyogenes*, mac = *macacae*, mut = *mutans*, udo = *pseudopneumoniae*

**Supplementary Figure 2.** (A) Sequence alignment (Geneious software) for selected orthologs of interest that substitute at least one critical PAM-contacting arginine residue within the region highlighted in red. A blue box marks the C-terminal component grafted onto truncated SpyCas9 to form dCas9 hybrids. (B) Table listing the homology shared within and outside of the box to those regions in the corresponding SpyCas9 and SmaCas9 reference sequences. (C) Chromatograms representing the PAM-SCANR based enrichment of variant-recognizing PAM sequences from a 5'-NNNNNNNNN-3' library.

A

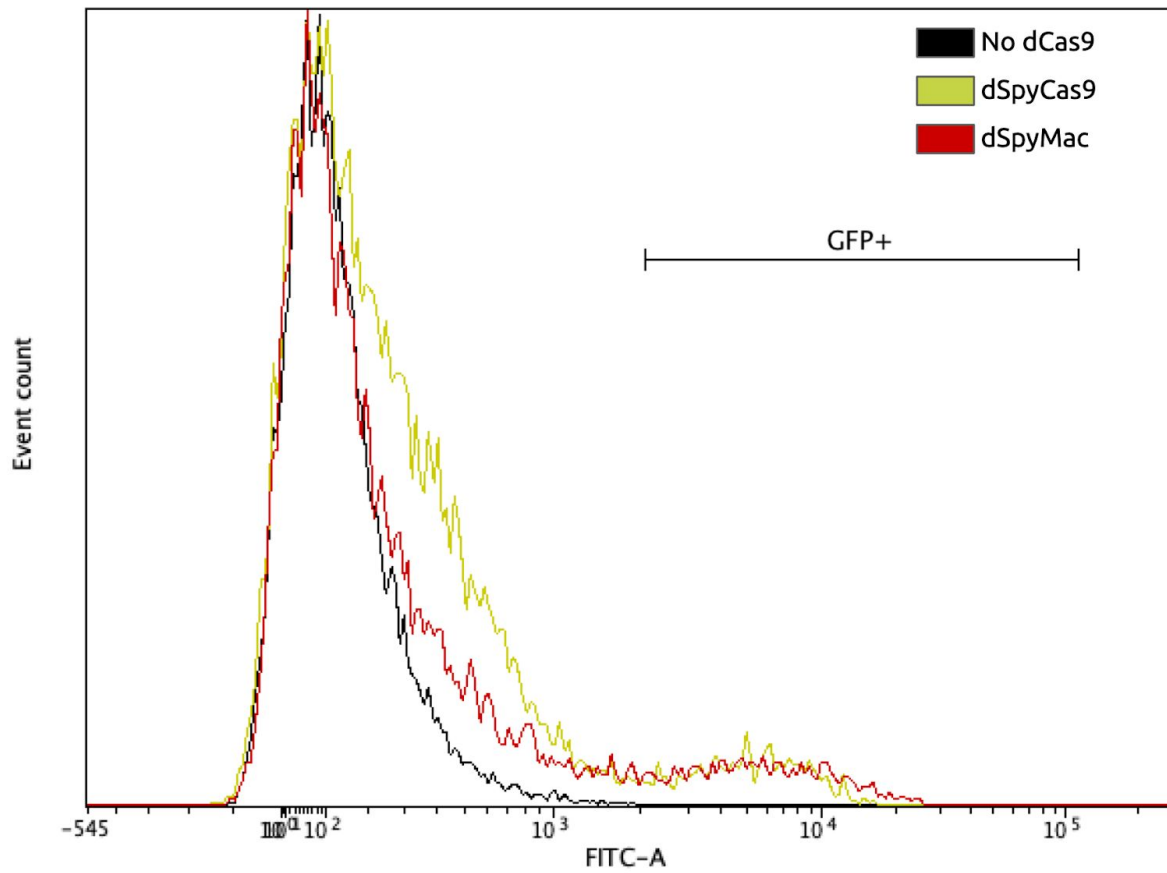
**Smac CRISPR Cassette (NCBI: NZ\_AEUW02000001)****Smut1 CRISPR Cassette (NCBI: AB742518)****Smut2 CRISPR Cassette (NCBI: NZ\_ALYX01000034)**

B

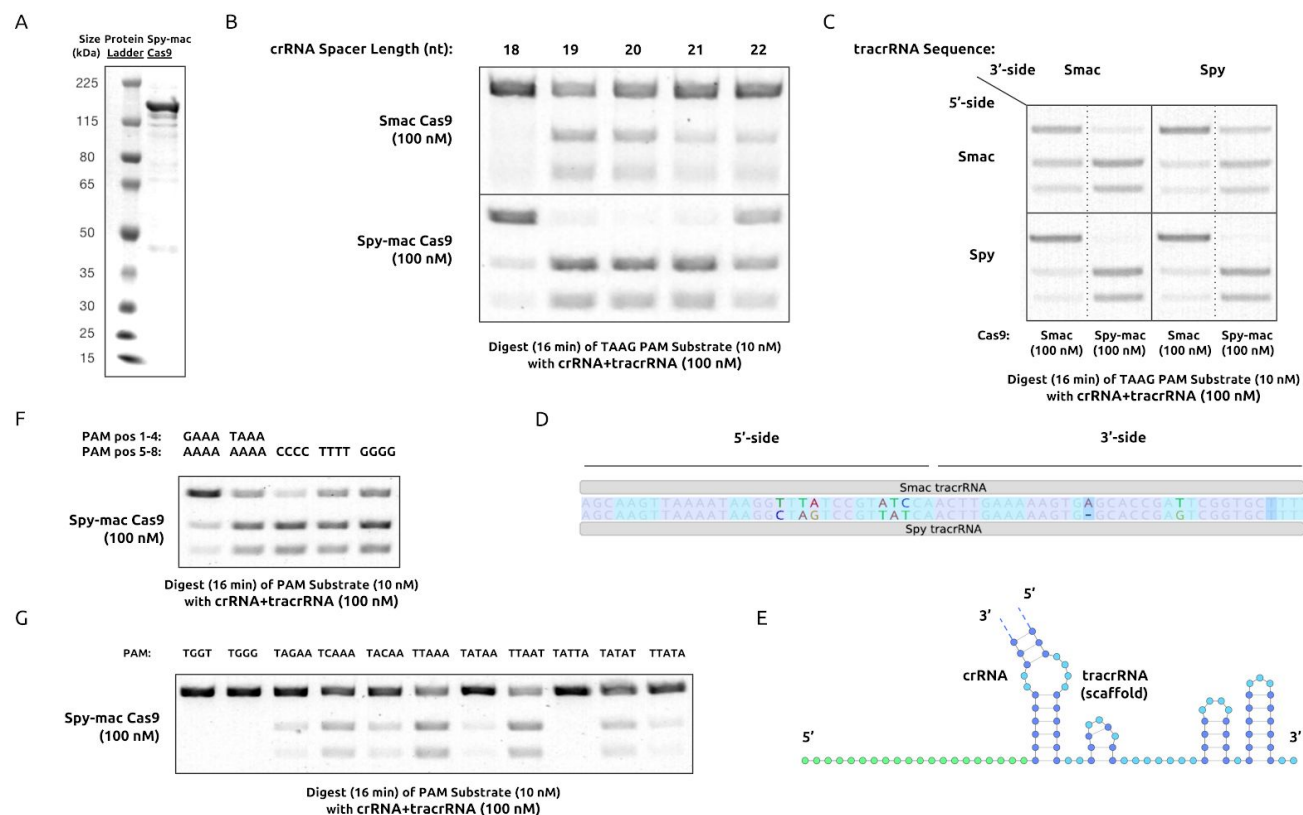
**Notable Spacer Mappings****Putative Protospacer PAM Strain/Species Source**

|                      |             |                                               |
|----------------------|-------------|-----------------------------------------------|
| CTAATCTAGCTAATGGTGA  | TC AATTTTTT | <i>Macaca fascicularis</i> (Host; Primate)    |
| AACATTTGAGATACCTCTTT | AAAGTAAGGC  | <i>Macaca fascicularis</i> (Host; Primate)    |
| GGC AAAAGTAGCCTGAAT  | CTAAGGATAG  | <i>Macaca fascicularis</i> (Host; Primate)    |
| TCCATGAAATTTCCAATCTG | TAAAGTCTGT  | <i>Streptococcus phage APCM81/M102/M102AD</i> |
| TACGTTCTTAAACTTGCCC  | AAATCC----  | <i>Parasteatoda tepidiorum</i> (House Spider) |
| GGGATGATTACGCTTGAGTT | AATAGAAATG  | <i>Candidatus Izimiplasma</i> (Bacteria)      |
| GATATTGACTGTGCTGTTTC | CAAATCTCT   | <i>Streptococcus phage M102AD</i>             |
| TTTTGAGTTTCTTACAAAC  | TAAAGTTTC   | <i>Streptococcus phage APCM81</i>             |
| TTTTGCTTTCTTACAAAT   | CAAGATTTC   | <i>Streptococcus phage M102/M102AD</i>        |
| TAAAGCTCTGCTGTGCT    | TTAAGGTGCC  | <i>Danio rerio</i> (Zebrafish)                |
| GATAAAGGAAAGCCATGAC  | TCACGCCAAA  | <i>Ovis canadensis</i> (Bighorn Sheep)        |
| GATAAAGGCAAGCCATGAC  | CAAGGCCATC  | <i>Streptococcus phage APCM81</i>             |
| IGTGCTTGCTGCACAGTAG  | AAGTTCTCT   | <i>Lepisosteus oculatus</i> (Spotted Gar)     |
| AAAAAGCAGGAGTAATCAG  | TAAAGATAC   | <i>Pieris rapae</i> (Cabbage White Butterfly) |
| AAAAAGCAGGIGTAATCAG  | CAAAACACTA  | <i>Streptococcus phage M102</i>               |
| ACTAAGACCGCAACTTAAC  | TAACTTGAA   | <i>Streptococcus phage M102/M102AD</i>        |
| ACGAAGACCGCAACTTAAC  | TAACTTAGAA  | <i>Streptococcus phage APCM81</i>             |
| GCTAAACCGTGACAGATGA  | TAACTTCGT   | <i>Streptococcus phage M102</i>               |
| GCTAAACCTGTCAGATGA   | TAAATTTGT   | <i>Streptococcus phage APCM81</i>             |
| TCAAAGTCTCTGAGTAGTCA | TATATGTAAA  | <i>Streptococcus phage M102/M102AD</i>        |
| TCAGAGTCTCTGAGTAGTCA | TATATGTAAA  | <i>Streptococcus phage APCM81</i>             |

**Supplementary Figure 3.** (A) Annotated CRISPR cassettes obtained from the genomes corresponding to orthologs that substitute both PAM-contacting arginine residues to glutamine. (B) Mappings of CRISPR cassette spacers to their putative target source for listed crRNA, identified via an online BLAST and/or SPAMALOT. SPAMALOT uncovered most cases of mismatch-tolerated mappings to *Streptococcus phage*. Underlined bases indicate mismatches that are tolerated for the mapping. Additional line spacing separates analysis for each CRISPR cassette.

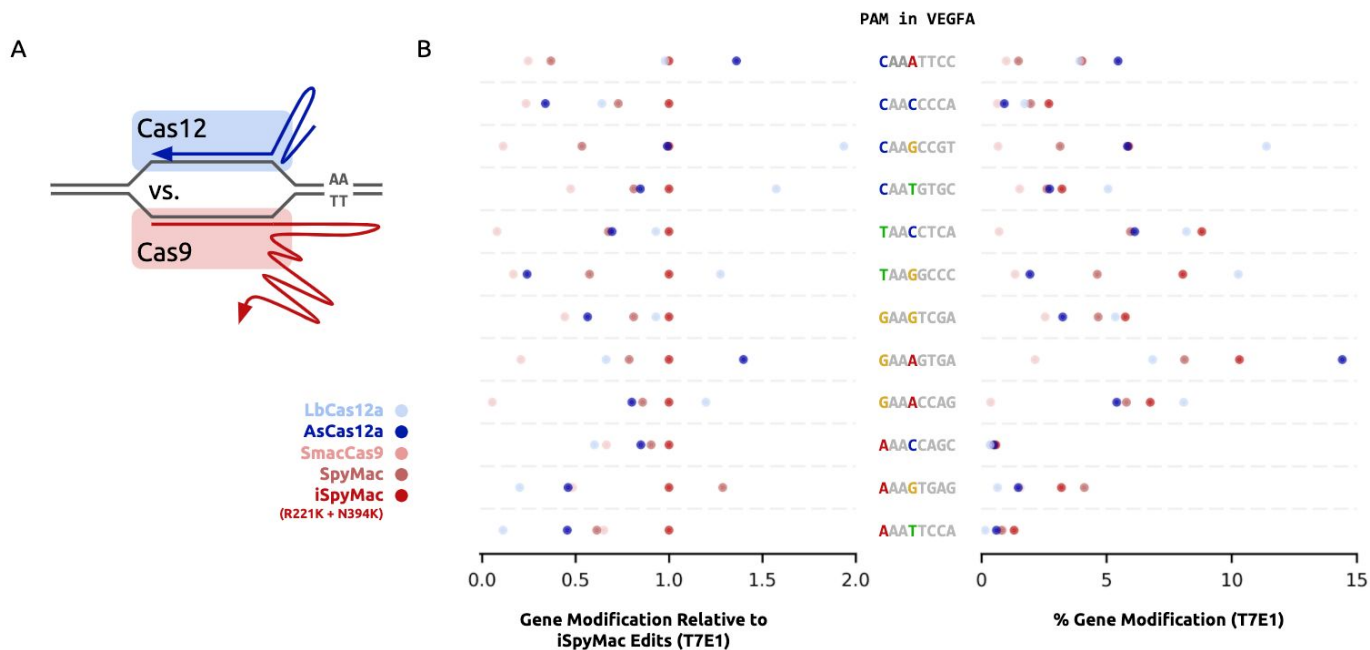


**Supplementary Figure 4.** 10,000 gated events for data analysis based on default FSC/SSC parameters for *E. coli*. The GFP+ gate was established both by a "no dCas9" negative control and a "dSpyCas9" positive control.



**Supplementary Figure 5.** (A) SDS-PAGE gel image of SpyMac after purification by affinity chromatography. SYBR-stained agarose gels running *in vitro* digestion reactions are shown that assay dependencies on (B) crRNA spacer length and (C) tracrRNA sequence origin. (D) Sequence alignment (Geneious software) of tracrRNA from *S. pyogenes* and *S. mutans* highlighted in a color code that reflects the base-pairing in their (E) duplex gRNA secondary structure. SYBR-stained agarose gels running *in vitro* digestion reactions are shown that assay dependencies on (F) positions 5-8 in the PAM sequence and (G) increments to the distribution of adenine content in positions 1-5 in the PAM sequence. All experiments in this figure were performed in independent biological duplicates (n=2), and experiments were repeated to ensure reproducibility.





**Supplementary Figure 6.** (A) Schematic diagram for matching Cas9 and Cas12a guides in a manner that enforces their recognition of the same PAM sequence and therefore facilitates their comparison (a "Cas12a vs Cas9 Comparator"). (B) Dot plots of absolute and relative gene modification efficiency in HEK293T cells by Cas9 and Cas12a variants targeting common PAM sequences located in the *VEGFA* gene. Values were quantified in a T7E1-based assay and are consistent with two independent biological replicates (n=2) that were run in parallel. Source data are provided as a Source Data file.