A scoutRNA is required for some type V CRISPR-Cas systems

Lucas B. Harrington[1,13,14], Enbo Ma[1,14], Janice S. Chen[1,13], Isaac P. Witte[1], Dov Gertz[10], David Paez-Espino[11,12], Basem Al-Shayeb[5], Nikos C. Kyrpides[11,12], David Burstein[4,10], Jillian F. Banfield[4,6], Jennifer A. Doudna[1-3,6-9*]

[1]Department of Molecular and Cell Biology; [2]California Institute for Quantitative Biosciences; [3]Department of Chemistry; [4]Department of Earth and Planetary Science; [5]Department of Plant and Microbial Biology; [6]Innovative Genomics Institute; [7]Howard Hughes Medical Institute; University of California, Berkeley, California 94720, USA. [8]Molecular Biophysics & Integrated Bioimaging Division, Lawrence Berkeley National Laboratory. [9]Gladstone Institutes, University of California, San Francisco, California 94114. [10]School of Molecular Cell Biology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel. [11]Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [12]Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

[13]Present address: Mammoth Biosciences Inc., South San Francisco, CA

[14]Equal contributions

*Correspondence to: doudna@berkeley.edu

SUMMARY

CRISPR-Cas12c/d proteins share limited homology with Cas12a and Cas9 bacterial CRISPR RNA (crRNA)-guided nucleases used widely for genome editing and DNA detection. However, Cas12c (C2c3)- and Cas12d (CasY)-catalyzed DNA cleavage and genome editing activities have not been directly observed. We show here that a short-complementarity untranslated RNA (scoutRNA), together with crRNA, is required for Cas12c/d-catalyzed DNA cutting. The scoutRNA differs in secondary structure from previously described tracrRNAs used by CRISPR-Cas9 and some Cas12 enzymes, and in Cas12d-containing systems, scoutRNA includes a conserved five-nucleotide sequence that is essential for activity. In addition to supporting crRNA-directed DNA recognition, biochemical and cell-based experiments establish scoutRNA as an essential cofactor for Cas12c-catalyzed pre-crRNA maturation. These results define scoutRNA as a third type of transcript encoded by a subset of CRISPR-Cas genomic loci and explain how Cas12c/d systems avoid requirements for host factors including Ribonuclease III for bacterial RNA-mediated adaptive immunity.

INTRODUCTION

CRISPR-Cas (clustered regularly interspaced short palindromic repeats, CRISPR associated) systems provide bacteria and archaea with adaptive immunity against infectious agents (Barrangou et al., 2007). RNA-guided nucleases are central to these pathways, recognizing and cutting double-stranded DNA to trigger degradation of targeted sequences in phage and plasmids (reviewed in Marraffini, 2015; Wright et al., 2016). In addition, the Cas9 and Cas12a enzymes found within type II and type V CRISPR-Cas systems, respectively, are now widely used for genome editing applications in eukaryotic cells and organisms based on their programmable ability to trigger DNA repair at desired sites (reviewed in Knott and Doudna, 2018; Wu et al., 2018).

Two types of noncoding RNAs have been identified as central components of CRISPR-Cas systems, CRISPR RNA (crRNA) and transactivating CRISPR RNA (tracrRNA). CRISPR RNA is used by all known CRISPR systems, as it provides the sequence recognition capability of these pathways (Brouns et al., 2008). Produced by transcription and processing of the CRISPR sequence array, which includes direct repeats separated by target-derived spacers, crRNAs guide Cas nucleases to cut DNA with complementarity to a ~20-nucleotide crRNA segment (Bolotin et al., 2005; Brouns et al., 2008; Garneau et al., 2010; Hale et al., 2009; Mojica et al., 2005; Pourcel et al., 2005). A second RNA, tracrRNA, is encoded within type II and some type V CRISPR-Cas genomic loci, where it is necessary for both crRNA maturation (Deltcheva et al., 2011; Chylinski et al., 2013; Shmakov et al., 2015) and CRISPR-Cas9-mediated DNA cleavage (Jinek et al., 2012). Extended base pairing complementarity between tracrRNA and the direct-repeat segment of crRNA creates a double-stranded RNA structure that is a substrate for Ribonuclease III-catalyzed processing (Deltcheva et al., 2011). The resulting dual-RNA guide is required for CRISPR-Cas9-catalyzed double-stranded DNA recognition and cleavage (Jinek et al., 2012).

The identification of divergent CRISPR-Cas systems, particularly within metagenomic sequencing datasets, has revealed new enzymes with only limited sequence similarity to known proteins. Among these, the Cas12c and Cas12d enzymes (also known as C2c3 and CasY, respectively) have attracted interest due to their small size and, in the case of Cas12d, predominant occurrence within the compact genomes of Candidate Phyla Radiation (CPR) bacteria (Burstein et al., 2017; Shmakov et al. 2015). However, with the exception of DNA targeting activity detected indirectly for Cas12d (Burstein et al., 2017), Cas12c/d-catalyzed DNA cleavage has not been observed.

We wondered whether Cas12c/d enzymes require additional components, either encoded in the CRISPR-Cas locus or elsewhere in host genomes, for RNA-guided DNA cutting. Here we show that a third type of CRISPR-Cas-encoded RNA, a short-complementarity untranslated RNA (scoutRNA), assembles with Cas12c/d and crRNA to create a functional DNA-targeting complex. Transcriptomic sequencing data indicate that processing of an initial precursor transcript generates scoutRNA, which includes only a short but highly conserved 3-5 -nucleotide sequence that is complementary to the repeat sequence in the crRNA. Biochemical experiments reveal that scoutRNA binds directly to Cas12d, where it functions together with crRNA to enable site-specific double-stranded DNA cleavage. We also found that scoutRNA is required for pre-crRNA processing by Cas12c by a mechanism distinct from any known crRNA maturation mechanism. These findings explain why Cas12c/d CRISPR systems can exist in the compact genomes of CPR bacteria that lack the Ribonuclease III enzyme needed for tracrRNA-mediated crRNA processing. Together our results uncover a new category of CRISPR-Cas systems defined by a unique RNA component and activation mechanism, showing how diversification of these pathways could have assisted their spread among divergent microbial populations.

RESULTS

Cas12c/d represent compact CRISPR-Cas systems found in tiny genomes

Class 2 CRISPR-Cas systems typically include a single large (100-200 kDa) CRISPR-associated (Cas) protein that catalyzes RNA-guided cleavage of DNA or RNA substrates. Searches to identify new class 2 proteins in bacterial metagenomic datasets revealed the existence of proteins classified as Cas12d, defined by proximity to a CRISPR array and the conserved CRISPR-associated gene *cas1* (Burstein et al., 2017). Comparative sequence and protein architecture analysis showed that CasY (now known as Cas12d) proteins are most closely related to the CRISPR-C2c3 family of enzymes (renamed Cas12c); for simplicity we refer to this CRISPR-Cas subclass as Cas12c/d (Fig. 1A, B). These proteins belong to the CRISPR-Cas type V superfamily, enzymes that contain a single RuvC nuclease domain that, in other type V-family enzymes, is responsible for RNA-guided DNA cleavage.

We identified, based on comparative sequence analysis, 23 distinct variants of Cas12c/d proteins from microbial organisms populating diverse environments including hot springs, Antarctic sea ice and insect microbiomes (Supp. Fig. 1A). Notably, Cas12d genes and their CRISPR-Cas genomic loci occur primarily in the compact genomes of Candidate Phyla Radiation (CPR) bacteria, a microbial super-phylum characterized by small cell and genome sizes (Fig. 1B, C). Consistent with this phylogenetic distribution, Cas12c/d systems are streamlined relative to other type V CRISPR-Cas enzymes, frequently occurring in CRISPR-Cas operons lacking any other *cas* genes except for *cas1*, which encodes the CRISPR integrase (Yosef et al., 2012; Nuñez et al., 2015; Wright et al., 2019) (Fig. 1A; Supp. Fig. 1A, B).

Although initial results demonstrated indirectly that Cas12c and Cas12d are capable of RNA-guided DNA interference (Burstein et al., 2017; Yan et al., 2019), no direct RNA-programmed DNA targeting activity has been detected for Cas12c/d proteins. We hypothesized that these proteins require a short sequence in DNA known as the protospacer-adjacent motif (PAM) for recognition of RNA-guided double-stranded DNA. To test this possibility, we transformed *E. coli* expressing a minimal Cas12d locus with a dsDNA plasmid containing a randomized PAM region next to a sequence matching the target-encoding sequence (spacer) in the Cas12d CRISPR array. Depletion analysis of plasmids in resulting *E. coli* transformants revealed that Cas12d requires a T-enriched PAM sequence for DNA cleavage, similar to the PAM preference detected for other type V-family CRISPR-Cas enzymes (Fig. 1D, E). The Cas12d PAM is a minimal TR (R=A/G) sequence (Burstein et al., 2017; Chen et al., 2019) in contrast to the 4-nt PAM required for most Cas12a orthologs (Zetsche et al., 2015). This TR PAM allows for a ten-fold increase, relative to Cas12a proteins, in the number of targetable sites in recently published CPR bacteriophage genomes (Fig. 1F).

Cas 12c/d requires scoutRNA, a non-coding transcript necessary for DNA interference

Efforts to detect RNA-guided Cas12c/d-catalyzed DNA cleavage directly, or to reconstitute this activity biochemically, has proved elusive, raising the possibility of a missing component that is necessary for enzymatic activity. Inspection of multiple *cas12d*-containing genomic loci revealed the presence of a noncoding region between the CRISPR array and *cas12d* (Fig. 2A). To test the requirement for this noncoding sequence for Cas12d function, we conducted plasmid transformation experiments in *E. coli* in which the CRISPR-Cas12d locus was expressed with a plasmid-complementary crRNA and with or without the noncoding sequence in the locus (Fig. 2A). The results showed that plasmid transformation could only be prevented by crRNA-guided Cas12d targeting when the full-length noncoding sequence and a repeat sequence upstream of the spacer was present in the CRISPR-Cas12d locus (Fig. 2A).

Examination of Cas12c- and Cas12d-containing CRISPR-Cas genomic loci identified potential homologs of this non-coding sequence that in many cases includes a short conserved pyrimidine-rich sequence with base pairing complementarity to a short purine-rich sequence in the corresponding CRISPR array repeat (Supp. Fig. 2A-C). Northern blotting of RNA extracted from a Cas12c protein expressed in *E. coli* from its cloned native locus demonstrated the presence of the corresponding transcript of similar size to the *in vitro* transcribed transcript (Fig.

2B). Notably, this RNA was not detected when the corresponding genomic region was deleted from the expression plasmid or when an oligonucleotide probe with a sequence complementary to the opposite genomic strand was used. These results suggest conservation of this noncoding RNA between the Cas12c and Cas12d subtypes.

To examine the *in vivo* expression of the CRISPR-Cas12d locus, we sequenced the RNA isolated from affinity-purified Cas12d protein expressed in *E. coli* harboring a CRISPR-Cas12d locus-containing plasmid. In addition to transcripts corresponding to the CRISPR array, as expected, we found an abundant small RNA species produced from the noncoding sequence between the CRISPR array and the *cas12d* gene (Fig. 2C). This 50-100 nt RNA is transcribed in the same direction as the CRISPR array. Unlike *trans*-activating CRISPR RNA (tracrRNA), originally identified in type II CRISPR-Cas systems and required for pre-crRNA maturation (Deltcheva et al., 2011) and CRISPR-Cas9 cleavage activity (Jinek et al., 2012), this transcript bears only limited complementary to the repeat region of Cas12d crRNAs (Supp. Fig. 2C). Furthermore, its predicted secondary structure differs from tracrRNA and contains a short unpaired RNA segment that exposes the limited region of crRNA complementarity (Fig. 2C).

We next examined environmental metatranscriptomic data (Brown et al., 2015) to determine whether this RNA is also produced in native uncultured hosts of Cas12d. We found limited RNA reads mapping to the CRISPR array, likely due to array diversity not represented in the reference genome. However, a transcript analogous to the scoutRNA identified in *E. coli,* with similar secondary structure and limited complementarity to the CRISPR array repeat sequence, was observed (Fig. 2D). We noted that scoutRNA transcript boundaries detected in metatranscriptomic data were variable, perhaps reflecting differential RNA processing at transcript ends and in cells within a large population. As observed for tracrRNAs, variability between in-silico prediction and mature, processed transcripts could impact the ability to predict scout RNA sequences in other systems.

Reconstitution of a Cas12d-scoutRNA-crRNA DNA targeting complex

We next tested whether purified Cas12d is capable of crRNA-guided DNA cleavage in the presence of the scoutRNA. We incubated purified Cas12d-crRNA complexes with radiolabeled target oligonucleotides (ssDNA, dsDNA, and ssRNA) bearing 18-nucleotide sequence complementary to the crRNA guide sequence, in the absence or presence of scoutRNA, and analyzed these substrates for Cas12d-mediated cleavage. Cleavage products for a crRNA-complementary dsDNA were only observed in the presence of scoutRNA (Fig. 3A; Supp. Fig. 3B). However, no cleavage was observed for the Cas12c ortholog tested in this study under the current reaction conditions (Supp. Fig. 3C). Cleavage site mapping showed that like other type V-family CRISPR-Cas enzymes, Cas12d generates a staggered dsDNA cut with a ~9-nt overhang (Supp. Fig. 3D). These results establish the scoutRNA as a required component of Cas12d-catalyzed RNA-guided dsDNA cleavage.

Type V CRISPR-Cas systems have been shown to target ssDNA, dsDNA and ssRNA (Zetsche et al., 2015; Chen et al., 2018; Yan et al., 2019). Using the functionally reconstituted Cas12d, we investigated the substrate preferences of this complex (Fig. 3B; Supp. Fig. 3E,). We observed rapid and precise cleavage of both ssDNA and dsDNA substrates with base pairing complementarity to the Cas12d guide RNA sequence. In contrast, no detectable cleavage was observed for RNA. Following recognition of on-target substrates, many Type V proteins are activated as non-specific ssDNA endonucleases (Chen et al., 2018). We tested whether this activity is also a property of Cas12d by providing a dsDNA activator molecule matching the guide RNA sequence (Fig. 3C, Supp. Fig. 3F). Incubating this activated complex with non-specific ssDNA, dsDNA and ssRNA revealed that Cas12d displays robust *trans* ssDNA cutting activity; no such non-specific activity was detected for dsDNA or ssRNA substrates. We used this *trans* cleavage activity to further investigate the fidelity of Cas12d for its on target dsDNA substrate, using *trans* cleavage as a proxy for on target DNA binding and cleavage. By tiling mismatches across the target DNA, we observed a PAM-proximal seed region similar to other

Class 2 CRISPR effectors (Fig. 3D). Notably, the protein was sensitive to mismatches across the majority of the guide sequence, in contrast to Cas12a which shows a more focused seed region. Further analysis will be needed to compare more directly the fidelity of these systems. Together, these results establish Cas12d as a dual-guided, programmable DNA targeting nuclease.

Mechanism of crRNA recognition

Inspection of multiple different scoutRNA sequences identified a five-nucleotide sequence (5'-GCCUU-3') that is conserved in Cas12d-associated scoutRNAs (Supp. Fig. 2B) and predicted to occur within a secondary structural unpaired region (Fig. 4A; Supp. Fig. 2A). This sequence is complementary to a five-nucleotide sequence found in the CRISPR array repeat sequence and thus present in every crRNA transcript generated from the Cas12d arrays, suggesting a possible base-pairing interaction between scoutRNA and crRNA. Nitrocellulose filter binding experiments showed that co-existence of crRNA and scoutRNA bound with higher affinity to purified Cas12d protein than either RNA alone (Fig. 4B; Supp. Fig. 4B). We next tested mutated versions of scoutRNA bearing altered sequences in the conserved segment, and tested these in Cas12d-catalyzed dsDNA cleavage assays (Fig. 4C, D; Supp. Fig. 4A). We also tested crRNA bearing compensatory mutations designed to restore base pairing with the altered scoutRNAs (Fig. 4C). DNA cleavage results showed that scoutRNA mutations disrupted Cas12d-catalyzed DNA cleavage, and this disruption was not restored by creating compensatory mutations in the crRNA. These findings differ from those observed in analogous experiments with *S. pyogenes* Cas9, where tracrRNA sequence mutations had no effect on DNA cleavage efficiency when the compensatory mutation was made in the repeat (Fig. 4C).

These results suggest that unlike tracrRNA, which forms an extensive base pairing interaction with crRNA in type II CRISPR-Cas systems (Deltcheva et al., 2011; Chylinski et al., 2013), scoutRNA assembly with Cas12d and crRNA may involve only short sequence specific recognition of the conserved 5-nucleotide scoutRNA sequence. Our data neither confirm nor refute the hypothesis that scoutRNA forms a base-paired interaction with crRNA, since compensatory mutations that maintain this base pairing potential but alter the RNA sequence were defective or inactive for RNA-guided Cas12d activity (Fig. 4C). To test this further, we created a mutant scoutRNA that collapsed the predicted unpaired region containing the conserved 5-nts without altering the conserved sequence itself. No Cas12d-catalyzed RNA-guided dsDNA cleavage was detected in the presence of this modified scoutRNA (Fig. 4D; Supp. Fig. 4A). In contrast, mutations that maintain base pairing in the flanking regions of scoutRNA had no impact on cleavage rate (Fig. 4D). Together, these results support an essential role for the conserved 5-nt sequence in scoutRNA and suggest, but do not confirm, its formation of a base pairing interaction with a short complementary region of the crRNA.

A dual RNA-guided pre-crRNA autoprocessing mechanism

In bacteria, CRISPR transcripts are often generated as precursors that must be cleaved to produce the mature crRNAs that guide DNA recognition. Type II CRISPR systems comprising Cas9 use tracrRNA to create an extensive double-stranded structure with pre-crRNA for recognition and processing by Ribonuclease III (Chylinski et al., 2013; Deltcheva et al., 2011). In contrast, the Cas12a subfamily of type V CRISPR systems possesses internal ribonucleolytic activity for auto-cleavage of crRNA precursors (Fonfara et al., 2016). We wondered how crRNAs are produced in Cas12c/d systems, given that limited base-pairing complementarity between scoutRNA and crRNA might preclude association in the absence of a Cas12c/d protein. In addition, the genomes from which these systems are derived do not always harbor genes encoding Ribonuclease III, implying that another mechanism for crRNA production may be involved.

To test the possibility that Cas12c itself catalyzes pre-crRNA maturation, we generated a set of substrates designed to detect Cas12c-mediated pre-crRNA processing. Initial experiments in which cleavage was expected at a position in the repeat upstream of the spacer, analogous to the processing site in pre-crRNAs of Cas12a-type systems, resulted in no detectable cleavage product. However, we were surprised to observe robust scoutRNA-dependent processing of a pre-crRNA substrate that enabled detection of cutting at a position on the opposite end of the pre-crRNA (Fig. 5A), suggesting a cleavage mechanism distinct from that observed for other CRISPR-Cas enzymes known to process their own pre-crRNAs, including Cas12a or Cas13a (East-Seletsky et al., 2016; Fonfara et al., 2016).

We next mutated the regions upstream or downstream of the processed crRNA spacer sequence to determine the mechanism of substrate recognition. Mutation of the upstream repeat sequence resulted in complete ablation of the RNA processing activity on the downstream spacer, likely due to lack of binding to the scoutRNA. By comparison, mutation of the predicted cleavage site still supported pre-crRNA processing (Fig. 5B). These results suggest that the spacer is measured by a ruler mechanism whereby Cas12c recognizes the sequence of the upstream repeat and cleaves downstream from the recognition site 18 nt. away. This mechanism is distinct from Cas12a and Cas13a enzymes, which catalyze pre-crRNA cleavage at the recognized CRISPR repeat sequence. Mutations of the scoutRNA to alter the predicted secondary structure at or near the short conserved sequence had variable effects on the rate of pre-crRNA processing (Supp. Fig. 5A) and we did not observe conclusive pre-crRNA processing by Cas12d in the same reaction conditions (Supp. Fig. 5B). The disproportionate impact of scoutRNA mutations on Cas12d-mediated DNA cleavage compared to Cas12c-mediated pre-crRNA processing could reflect differences in enzyme catalytic activities, CRISPR-Cas system functionality or both.

Together, these results reveal a new mechanism of crRNA maturation that requires both the scoutRNA and Cas12c but not an external ribonuclease. Based on scoutRNA conservation, it is likely that this mechanism extends to the Cas12c/d family of enzymes and that scoutRNA-dependent pre-crRNA processing is an inherent activity of these proteins that may enable their propagation in organisms lacking Ribonuclease III and related activities.

DISCUSSION

CRISPR-Cas systems have evolved in diverse microbial populations to provide adaptive protection from bacteriophage infection and plasmid transformation. These systems have been shown to employ two kinds of non-coding RNA molecules, crRNA and tracrRNA. Whereas crRNA is used universally to identify foreign nucleic acids by base pairing, tracrRNA has been found only in type II and the Cas12b (C2c1) and CasX (Cas12e) type V CRISPR systems, where it functions both during pre-crRNA maturation and Cas9/Cas12b/CasX targeting complex assembly. We show in this study that Cas12c/d type V CRISPR-Cas systems encode and employ a distinct type of noncoding RNA, scoutRNA, which is required for both pre-crRNA maturation and Cas12c/d targeting complex formation. For the CRISPR-Cas12c/d genomic loci examined in this study, none were found to encode a tracrRNA and all encoded a scoutRNA, according to the criteria described here. Unlike tracrRNAs, scoutRNA sequences have minimal base pairing complementarity to the corresponding crRNA repeat sequence, and our data do not confirm the existence of base-pairing between scoutRNA and crRNA. The definition of the scoutRNA as distinct from tracrRNA also sets the stage for defining and naming CRISPR-Cas components according to their function rather than according to their order of discovery or proposed phylogenetic relationships.

In addition to a predicted secondary structure that precludes an extensive pre-crRNA base pairing interaction, the scoutRNA supports a mechanism of pre-crRNA processing that is distinct from those of either tracrRNA-dependent or independent processing systems. Instead of substrate recognition and cleavage occurring together in the tracrRNA-pre-crRNA duplex or pre-crRNA alone, scoutRNA supports Cas12c-catalyzed maturation by a mechanism in which substrate recognition and cleavage occur on separate segments of the pre-crRNA. This is notably inconsistent with Ribonuclease III-catalyzed RNA processing, which involves double-stranded RNA recognition and cutting that generates 2-nt. 3' overhangs in the cleavage product (Court et al., 2013; Nicholson, 2014). This difference in pre-crRNA processing mechanisms supports the conclusion that scoutRNA is functionally distinct from tracrRNAs as originally defined (Deltcheva et al., 2011; Chylinski et al., 2013).

Until now, CRISPR-Cas systems have been categorized according to their protein components, and phylogenetic relationships are derived from protein homologies. The existence of scoutRNA suggests a new possibility for categorization based on noncoding RNA composition. Three RNA-based classes of CRISPR-Cas systems include those using crRNA and tracrRNA, those using crRNA alone, and those using crRNA and scoutRNA (Fig. 6). The role of a conserved five-nucleotide crRNA-complementary segment in some scoutRNAs suggests a possible direct base pairing interaction with crRNA that would presumably occur only within the context of the Cas12c/d protein. The possible short segment of scoutRNA-crRNA base pairing is reminiscent of the short RNA-RNA base pairing that occurs between snRNAs, forming the interactions required for association with proteins to form snRNPs. It remains to be determined how scoutRNA creates a stable interaction with crRNA and whether, like tracrRNA, it creates a structural scaffold for Cas protein assembly and conformational dynamics.

The unique properties of scoutRNA, including variable length and sequence diversity, offer possibilities for engineering that include creation of shorter forms that retain function, and possibly fusions with crRNA to form an sgRNA-type construct. These possibilities, combined with the minimal PAM required for DNA target recognition, could enhance Cas12c/d functionality for genome editing by providing ways to induce cellular delivery or append RNA-encoded capabilities. Continued exploration of scoutRNA diversity should reveal whether its detection can signal the presence of new CRISPR-Cas systems or protein variants that have yet to be identified.

REFERENCES

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Bolotin, A., Quinquis, B., Sorokin, A., and Ehrlich, S.D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. Microbiology *151*, 2551–2561.

Brouns, S.J.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J.H., Snijders, A.P.L., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960–964.

Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A., Wilkins, M.J., Wrighton, K.C., Williams, K.H., and Banfield, J.F. (2015). Unusual biology across a group comprising more than 15% of domain Bacteria. Nature *523*, 208–211.

Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A., and Banfield, J.F. (2017). New CRISPR-Cas systems from uncultivated microbes. Nature *542*, 237–241.

Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X., Palefsky, J.M., and Doudna, J.A. (2018). CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. Science *360*, 436–439.

Chen, L.-X., Al-Shayeb, B., Méheust, R., Li, W.-J., Doudna, J.A., and Banfield, J.F. (2019). Candidate Phyla Radiation Roizmanbacteria From Hot Springs Have Novel and Unexpectedly Abundant CRISPR-Cas Systems. Front. Microbiol. *10*, 928.

Chylinski, K., Le Rhun, A., and Charpentier, E. (2013). The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. RNA Biol. *10*, 726–737.

Court, D.L., Gan, J., Liang, Y.-H., Shaw, G.X., Tropea, J.E., Costantino, N., Waugh, D.S., and Ji, X. (2013). RNase III: Genetics and function; structure and mechanism. Annu. Rev. Genet. *47*, 405–431.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

East-Seletsky, A., O'Connell, M.R., Knight, S.C., Burstein, D., Cate, J.H.D., Tjian, R., and Doudna, J.A. (2016). Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. Nature *538*, 270–273.

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., and Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature *532*, 517–521.

Garneau, J.E., Dupuis, M.-È., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–71.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-guided RNA cleavage by a CRISPR RNA-Cas protein complex. Cell *139*, 945–956.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816–821.

Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. Science *361*, 866–869.

Marraffini, L.A. (2015). CRISPR-Cas immunity in prokaryotes. Nature *526*, 55–61.

Markowitz, Victor M; Chen I-Min A; Palaniappan Krishna; Chu Ken; Szeto Ernest; Grechkin Yuri; Ratner Anna; Jacob Biju; Huang Jinghua; Williams Peter; Huntemann Marcel; Anderson Iain; Mavromatis Konstantinos; Ivanova Natalia N; Kyrpides Nikos C (Jan 2012). "IMG: the integrated microbial genomes database and comparative analysis system"*.* Nucleic Acids Res. 40, D115–22.

Mojica, F.J.M., Díez-Villaseñor, C., García-Martínez, J., and Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. J. Mol. Evol. *60*, 174–182.

Nicholson, A.W. (2014). Ribonuclease III mechanisms of double-stranded RNA cleavage. Wiley Interdiscip. Rev. RNA *5*, 31–48.

Nuñez, J.K., Lee, A.S.Y., Engelman, A., and Doudna, J.A. (2015). Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. Nature *519*, 193–198.

Pourcel, C., Salvignol, G., and Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. Microbiology *151*, 653–663.

Wright, A.V., Nuñez, J.K., and Doudna, J.A. (2016). Biology and Applications of CRISPR Systems: Harnessing Nature’s Toolbox for Genome Engineering. Cell *164*, 29–44.

Wright, A.V., Wang, J.Y., Burstein, D., Harrington, L.B., Paez-Espino, D., Kyrpides, N.C., Iavarone, A.T., Banfield, J.F., and Doudna, J.A. (2019). A Functional Mini-Integrase in a Two-

Protein-type V-C CRISPR System. Mol. Cell *73*, 727–737.e3.

Wu, W.Y., Lebbink, J.H.G., Kanaar, R., Geijsen, N., and van der Oost, J. (2018). Genome editing by natural and engineered CRISPR-associated nucleases. Nat. Chem. Biol. *14*, 642–651.

Yan, W.X., Hunnewell, P., Alfonse, L.E., Carte, J.M., Keston-Smith, E., Sothiselvam, S., Garrity, A.J., Chong, S., Makarova, K.S., Koonin, E.V., et al. (2019). Functionally diverse type V CRISPR-Cas systems. Science *363*, 88–91.

Yosef, I., Goren, M.G., and Qimron, U. (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. *40*, 5569–5576.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., van der Oost, J., Regev, A., et al. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell *163*, 759–771.

ACKNOWLEDGMENTS

The authors thank members of the Doudna and Banfield laboratories for helpful comments on the manuscript. JAD acknowledges support for this project from the Centers for Excellence in Genomic Science of the National Institutes of Health (NIH) under award number RM1HG009490, and the National Science Foundation under award number 1817593. JAD is an Investigator of the Howard Hughes Medical Institute and is a Paul Allen Distinguished Investigator. JAD and JFB receive funding from the Somatic Cell Genome Editing Program of the Common Fund of the NIH under award number U01AI142817-02 and JFB acknowledges support from the NIH under award RAI092531A and from the Innovative Genomics Institute. DG was supported by an M.Sc. fellowship from the Edmond J. Safra Center for Bioinformatics at Tel Aviv University.

AUTHOR CONTRIBUTIONS

LBH, JSC, JCC, IPW, EM developed the project idea, conducted experiments, and prepared the manuscript and figures. DB, DP, DG, JFB and NK conducted and designed computational biology work and reviewed the manuscript. JAD provided financial support, assisted in experimental design and wrote the manuscript.

DECLARATION OF INTERESTS

JAD is a co-founder of Caribou Biosciences, Editas Medicine, Intellia Therapeutics, Scribe Therapeutics and Mammoth Biosciences. JAD is a scientific advisory board member of Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Synthego, Inari and Felix Biotechnology. JAD is a Director at Johnson & Johnson and has research projects sponsored by Biogen and Pfizer. UC Regents have filed patents related to this work on which DB, JFB, LBH, DP-E, JSC and JAD are inventors. LBH, JSC and

JAD are co-founders of Mammoth Biosciences. IPW served as a consultant for Mammoth Biosciences. JFB is a co-founder of Metagenomi.

FIGURE LEGENDS

Figure 1 | Cas12c/d are part of compact CRISPR systems found in tiny genomes.

A) Diagram of Type V-C and Type V-D CRISPR-Cas loci. Cas12c (C2c3) and Cas12d (CasY) that share minimal sequence similarity with Cas12a (Cpf1) except for the RuvC catalytic domain. B) Unrooted phylogentic tree showing Cas12c and Cas12d representatives. Newly identified orthologs are highlighted with colored circles (orange, Cas12c; blue, Cas12d) and greyed out circles mark previously described orthologs. Orthlogs used for experiment's in this study are identified by name. C) Host assignment for all CRISPR systems, Cas12c and Cas12d illustrating that Cas12d is highly enriched in Candidate Phyla Radiation (CPR) bacteria. D) A plasmid depletion screen for PAM-dependent inhibition of plasmid transformation showing that only target sequences adjacent to a TR sequence were efficiently depleted. E) Plasmid interference against individual PAM targets showing clearance of plasmids containing a TA or TG adjacent to the targeted sequence. F) Predicted number of sites in a CPR-associated bacteriophage genome that are targetable by Cas12a, Cas12e and Cas12d.

Figure 2 | Cas 12c/d requires a new kind of tracrRNA for DNA interference.

A) Plasmid transformation assay testing RNA-guided DNA targeting by CRISPR-Cas systems expressed in *E. coli*. Deletions were made of non-coding regions of the CRISPR locus and resulting plasmid transformation efficiencies are shown. B) Diagram of CRISPR-Cas12c genomic loci indicating a noncoding sequence between the *cas1* and *cas12c* genes; Northern blot using a radiolabeled DNA oligonucleotide probe (represented by red arrow) and affinity-purified samples of Cas12c when co-expressed with noncoding regions of the CRISPR locus, (IVT, *in-vitro* transcribed; KO, knockout). C, D) RNA-sequencing data corresponding to the

CRISPR-Cas non-coding locus, from samples that were affinity purified from *E. coli* expression (C) or obtained from metatranscriptomic analysis (D). Black diamonds in CRISPR loci cartoons represent repeats and white rectangles represent spacers. Purple rectangles correspond to the non-coding region and the predicted secondary structure of this region is shown to the right. Color scale represents base-pair probabilities.

Figure 3 | *cis*- and *trans*-cleavage activities of Cas12d Cas12d-catalyzed and crRNA-targeted DNA cleavage.

A) **ScoutRNA is essential for Cas12d-mediated dsDNA cleavage. In this assay, nontarget strand is 5'-end labeled, and the** reactions were conducted in the absence (-) or presence (+) of scoutRNA. B) Time course plots of *cis*-cleavage activity of Cas12d. C) Time course plots of *trans*-cleavage activity of Cas12d. The substrates of dsDNA, ssDNA and ssRNA used in this assay are non-specific to Cas12d crRNA. D) Cas12d cleavage activities on mutated dsDNA targets. In this assay, pairs of mismatched base pairs were tiled across the crRNA-target DNA strand duplex, and the resulting extent of crRNA-guided Cas12d-catalyzed dsDNA cleavage is shown.

Figure 4 | A short conserved sequence in scoutRNA is required for dsDNA targeting.

A) Cas12d-associated crRNA repeat sequence alignment. Conserved sequences are shown in black; predicted scoutRNA secondary structure and possible short base paired interaction between scoutRNA and crRNA repeat are also shown. B) Cas12d strongly binds to the complex from scoutRNA and crRNA. Data are from nitrocellulose filter binding assays with radiolabeled crRNA and/or scoutRNA as a function of Cas12d protein concentration; (*) indicates radiolabeled species when two RNAs were present in the binding reaction. C) The effect of reciprocal changes in guide RNA stem on Cas12d-mediated dsDNA cleavage. wt= wild-type and mut=mutation. D) Importance of 5 conserved nucleotides in Cas12d scoutRNA.

Mutants #4 and #5 contained sequence changes that maintained base pairing complementarity in the regions shown; mutant #2 contained nucleotide changes to create a complementary sequence on the strand opposite the conserved 5 nt. sequence.

Figure 5 | An RNase III-independent dual RNA-guided pre-crRNA processing mechanism. A) Timecourses of pre-crRNA cleavage in the presence or absence of purified Cas12c and scoutRNA, using a 5'-end radiolabeled 58-nt. pre-crRNA. B) Kinetics of scoutRNA-dependent Cas12c-catalyzed pre-crRNA cleavage using the pre-crRNA substrates shown.

Figure 6 | Three different types of RNA-guided CRISPR-Cas families defined by RNA components.

Non-coding RNAs enable functional classification of CRISPR-Cas enzymes into three distinct categories. All use crRNA, whereas a subset use either a canonical trans-activating CRISPR RNA (tracrRNA) and another subset use a short-complementarity untranslated RNA (scoutRNA).

STAR METHODS

Key Resources Table

REAGENT or RESOURCE	SOURCE
DNA Oligos	IDT
RNA Oligos	Synthego

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jennifer A. Doudna (doudna@berkeley.edu).

Experimental Model and Subject Details

N/A.

Phylogenetic Analysis

Amino acid sequences of proteins previously identified and new orthologs described in this manuscript were aligned using MAFFT and phylogenetic trees were constructed using RAxML. Trees were visualized using FigTree 1.4.4.

PAM depletion and plasmid interference

PAM depletion and plasmid interference assays were conducted as previously described (Burstein et al. 2017). Expression plasmids containing the native contig and non-coding sections (https://benchling.com/s/seq-c4cx5V2kzCCOLGsLplyY) were transformed into BL21(DE3*) E. coli.* After selection, these cells were grown to OD600=0.4 before pelleting and washing 3 times with ice cold 10% glycerol. The resulting cells were transformed 200ng of a plasmid library containing a randomized 7-nt section upstream of the region matching the spacer. After transformation the resulting cells were plated on selective medium containing carbenicillin and chloramphenicol for ~36hrs at room temperature. For plasmid interference assays the same procedure was followed but clonal plasmids were used in place of the randomized libraries. Serial dilutions of the electroporated cells were serially diluted and CFUs were counted.

Northern blotting

Both RNAs extracted from a Cas12c protein expressed in *E. coli* from its cloned native locus and transcribed in vitro were separated on 10% UREA-PAGE at 1 watt in 0.5X TBE after denatured in denature buffer of 95% of formamide, 0.001% bromophenol blue and 0.001% of xylene cyanol. The separated RNAs were blotted onto nylon membrane via semi-dry electroblotting in 0.5X TBE at 20 volts for 2 hours. The RNA blot was cross-linked in UV-cross linker and then pre-incubated for 3 hours at 45°C in hybridization buffer (40% formamide, 5X SSC, 3X Denhardt's, 200 ug/ml of salmon sperm DNA, and 0.1% SDS). The pre-incubated RNA blot was further incubated at 45°C overnight with 5'-end labeled DNA oligo in hybridization buffer. The blot was then washed once with 4X SCC, followed by 3 times with 0.1X SSC. The hybridization signals were detected and analyzed with Amersham Typhoon and ImageQuant (GE Healthcare).

Small RNA sequencing

RNAseq was conducted as previously described with modification (Minnier et al., 2018). Cells transformed with the native expression plasmid were grown in SOB to saturation overnight at 30°C. The resulting bacterial cell pellet was lysed by treatment with Lysozyme, SDS and hot phenol extraction. To prepare the RNA for sequencing it was treated with trubo DNase, rSAP and T4 PNK before inputting into the NEBnext small RNA sequencing illumine library kit. Resulting reads were trimmed with Cutadapt (Martin, 2018) and mapped using Bowtie 2 (Langmead et al., 2018).

Protein expression and purification

Cas12d (CasY) and Cas12c proteins were expressed in a modified pET vector containing an N-terminal 10×His-tag, maltose-binding protein (MBP) and TEV protease cleavage site. Proteins were purified as described elsewhere **(**Chen et al., 2018**)**, with the following modifications: *E. coli* BL21(DE3) containing Cas12d expression plasmids were grown in Terrific Broth at 16°C for 14 hr. Cells were harvested and re-suspended in lysis buffer (50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 5% (v/v) glycerol, 1 mM TCEP, 1 tablet of protease inhibitor/50 ml (Sigma-Aldrich)), disrupted by sonication, and purified using Ni-NTA resin. After overnight TEV cleavage at 4°C, proteins were purified over an OrthoTrap HP column, the elutes were further purified through a HiTrap Heparin HP column for cation exchange chromatography. The final gel filtration step (Superdex 200) was carried out in elution buffer containing 20 mM Tris-HCl, pH 7.5, 200 mM NaCl, 5% (v/v) glycerol and 1 mM TCEP. Purified Cas12d is shown in Supp. Figure 3A.

Nucleic acid preparation

DNA oligos were synthesized commercially (IDT, Integrated DNA Technologies, Inc., San Diego, CA USA), and PAGE-purified in-house before being radiolabeled for cleavage assays.

For generation of scout RNAs, the commercially synthesized T7-promoter-tagged DNA oligos served as templates for *in vitro* transcription reactions, which were performed as described elsewhere (Chen et al., 2018). crRNAs were commercially synthesized by IDT and PAGE-purified in-house. All DNA and RNA substrates are listed below.

DNA cleavage assays

Generally, Cas12d-mediated cleavage assays were carried out in cleavage buffer consisting of 20 mM Tris (pH 7.5), 100 mM NaCl, 10 mM $MgCl_2$, 1% glycerol and 0.5 mM DTT. For radiolabeled cleavage assays, the substrates of either target strand or non-target strand were 5′-end-labeled with T4 PNK (NEB, New England Biolabs) in the presence of gamma ^{32}P-ATP in 30 μl reactions. To form dsDNA substrates, the labeled substrate was annealed with excess cold target or non-target strand according to the labeled strand. In a typical Cas12d cleavage reaction, the concentrations of Cas12d, guide RNA and ^{32}P-labeled substrates were 100 nM, 120 nM and 2-4 nM, respectively. To carry out the assay, Cas12d was first incubated with its guide RNA(s) at room temperature for 15 min before addition of the labeled substrates at 37°C. Reactions were incubated for certain periods (min) of time as indicated and quenched with formamide-containing loading buffer (final concentration 45% formamide and 15 mM EDTA, with trace amount of xylene cyanol and bromophenol blue) for 3 min at 90°C. The reaction products were resolved by 12% urea-denaturing PAGE gel and quantified with Amersham Typhoon (GE Healthcare). The fraction of DNA cleaved at each time point was plotted as a function of time, and these data were fit with a single exponential decay curve using Prism 6 (GraphPad Software, Inc.), according to the equation: Fraction cleaved = $A \times (1 - \exp(-k \times t))$, where A is the amplitude of the curve, k is the first-order rate constant and t is time. All experiments were carried out at least in triplicate, with representative replicates shown in the figure panels.

For *trans*-cleavage assays, the Cas12d was first incubated with guide RNA(s) at room temperature for 15 min, then further incubated for another 15 min with activator at room temperature before addition of labeled substrates that are unrelated to guide RNA(s). The cleaved products were separated and quantified similarly as stated above.

Filter Binding Assays

Filter binding reaction was carried out in 30 ul reaction in filter-binding buffer (20 mM Tris [pH 7.5], 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol, 0.01% Igepal CA-630, 10 μg/ml yeast tRNA, and 10 μg/ml BSA). 1.2× concentration of Cas12d protein to unlabeled RNA was incubated with radiolabeled RNA (< 0.05nM) for 1 hr at room temperature. Tufryn, Protran, and Hybond-N+ were assembled onto a dot-blot apparatus in the order of Tufryn, Protran, and Hybond-N+ (from top to bottom). The membranes were washed twice with 50 μl equilibration buffer (20 mM Tris [pH 7.5], 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 5% glycerol) before the sample was applied to the membranes. Membranes were again washed twice with 50 μl equilibration buffer, air-dried, and visualized by phosphorimaging. Data were quantified with ImageQuant TL Software (GE Healthcare) and fit to a binding isotherm using Prism (GraphPad Software). Dissociation constants (K_D) is reported in the figure legends.

Cas12c pre-crRNA autoprocessing experiments

Processing reactions (total volume of 100 uL) contained 100 nM Cas12c, 120 nM scoutRNA, 3 nM 5′ radiolabeled pre-crRNA (wildtype, 3′ mutant, or 5′ mutant), and 1X Cleavage Buffer (20 mM Tris-HCl pH 7.5, 150 mM Kcl, 5 mM MgCl2, 1 mM TCEP). Prior to the addition of Cas12c to the reaction, scoutRNA and pre-crRNA were annealed in 1X Cleavage Buffer by incubating at 70ºC for 5 min followed by −2ºC/min to 25ºC. To test which components were essential for autoprocessing, Cas12c and scoutRNA were omitted from the reactions as indicated in Figure 5A. Reactions were incubated at 37ºC, and 15 uL of each reaction were quenched with 2x Quench Buffer (90% formamide, 25 mM EDTA, and trace bromophenol blue) at 0, 1, 5, 15, 30, and 60 min. Quenched reactions were heated to 95ºC for 2 min and run on a 15% denaturing polyacrylamide gel (7M Urea, 0.5xTBE). Products were visualized by phosphorimaging and band intensities were quantified using ImageQuant software.

DNA and RNA sequences

DNA substrates for cleavage assays:

Non-target (NT) GCCTGCCCGCAGACTAatcaataccaaactctggCGGCGTAAACTTTCCAGTC
Target (T) GACTGGAAAGTTTACGCCGccagagtttggtattgatTAGTCTGCGGGCAGGC

Used in trans-cleavage assays:
GACGACAAAACTTTAGATCGTTACGCTAACTATGAGGGCTGTCTGTGGAATGCTA

crRNAs used in this study:

RNA_382	ACCCGUAAAGCAGAGCGAUGAAGGCaUcaaUaccaaacUcUgg
RNA_386	GCGAUGAAGGCaUcaaUaccaaacUcUgg
RNA_387	GCGAUGAAGGCaUcaaUaccaaacUcUg
RNA_391	GCGAUGGGCGUaUcaaUaccaaacUcUgg

sccoutRNA:
RNA_396 CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUGGCCUUCUCCC UUAACCUAUGCCACUAAUGAUU

scoutRNAs of wild-type and mutations used in reciprocal mutation studies:

396-w.t.	CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUG**GCCUU**CUCCC UUAACCUAUGCCACUAAUGAUU
396-full mut	CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUG**ACGCC**CUCCC UUAACCUAUGCCACUAAUGAUU
396-mut1	CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUG**A**CCUUCUCCC UUAACCUAUGCCACUAAUGAUU
396-mut2	CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUGG**G**CUUCUCCC UUAACCUAUGCCACUAAUGAUU
396-mut3	CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUGGC**G**UUCUCCC UUAACCUAUGCCACUAAUGAUU
396-mut4	CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUGGCC**A**UCUCCC UUAACCUAUGCCACUAAUGAUU
396-mut5	CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUGGCCU**A**CUCCC UUAACCUAUGCCACUAAUGAUU

crRNAs of wild-type and mutations used in reciprocal mutation studies:

386-w.t.	GCGAUG**AAGGC**aUcaaUaccaaacUcUgg
386-full mut	GCGAUG**GGCGU**aUcaaUaccaaacUcUgg
386-mut1	GCGAUGAAGG**U**aUcaaUaccaaacUcUgg
386-mut2	GCGAUGAAG**C**CaUcaaUaccaaacUcUgg
386-mut3	GCGAUGAA**C**GCaUcaaUaccaaacUcUgg
386-mut4	GCGAUGA**U**GGCaUcaaUaccaaacUcUgg
386-mut5	GCGAUG**U**AGGCaUcaaUaccaaacUcUgg

RNA used for Casd12c (C2c3) RNA processing:
C2C3_1 Scout (143.1)
ggaUaccacccgUgcaUUUcUggaUcaaUgaUccgUaccUcaaUgUccgggcgcgcagcUagagcgaccU gaaaUcU
C2c3 RSRS (147)

ggagcaggaUUcaggUUgggUUUgaggAUCAAUACCAAACUCUGagcaggaUUcaggUUgggUUUgaggGAGACCacgcaGGUCUC

Casd12d scoutRNAs of wild-type (#1) and mutations:

#1 **CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUGGCCUUCUCCCUUAACCUAUGCC**

#2 **CUUAGUUAAGGAGAAGGCCAGGUUCUUUCGGGAGCCUUGGCCUUCUCCCUUAACCUAUGCC**

#3 **CUUAGUUAAGGAUGUUUCCAGGUUCUUUCGGGAGCCUUGGCCUUCUCCCUUAACCUAUGCC**

#4 **CUUAGUGCUGGAUGUUCCAGGUUCUUUCGGGAGCCUUGGCCUUCUCCCAGCACCUAUGCC**

#5 **CUUAGUUAAGGAUGUUCCAGGCGAUUUCGGUCGCCUUGGCCUUCUCCCUUAACCUAUGCC**

#6 **CUUAGUUAAGGAUGUUCCAGGUUCUUUCGGGAGCCUUGGCUUUCUCCCUUAACCUAUGCC**

Casd12c scoutRNAs of wild-type (#1) and mutations:

#1 **GGAUACCACCCGUGCAUUUCUGGAUCAAUGAUCCGUACCUCAAUGUCCGGGCGCGCAGCUAGAGCGACCUG**

#2 **GGAUACCACCCGUGCAUUGAGGUAUGGAUCAAUGAUCCGUACCUCAAUGUCCGGGCGCGCAGCUAGAGCGACCUG**

#3 **GGAUACCACCCGUGCAUUUUUUUCUGGAUCAAUGAUCCGUACCUCAAUGUCCGGGCGCGCAGCUAGAGCGACCUG**

#4 **GGAUACCACCCGUGCAUUUCUGGAUCAAUGAUCCGUUCUUCAAUGUCCGGGCGCGCAGCUAGAGCGACCUG**

#5 **GGAUACCACCCGUGCAUUUCUGACUCAAUGAGUCGUACCUCAAUGUCCGGGCGCGCAGCUAGAGCGACCUG**

#6 **GGAUACCACCCGUGGGAUUCUGGAUCAAUGAUCCGUACCUCAUCCUCCGGGCGCGCAGCUAGAGCGACCUG**

#7 **GGAUACCACCCGUGCAUUAAUGGAUCAAUGAUCCGUACCUCAAUGUCCGGGCGCGCAGCUAGAGCGACCUG**

Quantification and Statistical Analysis

Amino acid sequences of proteins previously identified and new orthologs described in this manuscript were aligned using MAFFT and phylogenetic trees were constructed using RAxML. Trees were visualized using FigTree 1.4.4. Products/images from both DNA/RNA cleavage assays and filter binding assays were visualized by phosphorimaging and band/dots intensities were quantified using ImageQuant software. Graphs from these data were generated from GraphPad Prism.

SUPPLEMENTAL FIGURE LEGENDS

Supplemental Figure 1 | Cas12c/d are part of compact CRISPR systems found in tiny genomes.

A) Schematics showing the locus composition of Cas12d (CasY) and Cas12c (C2c3) CRISPR-Cas systems identified in metagenomic datasets and in the Integrated Microbial Genomes database (Chen et al., 2019; Markowitz et al., 2012). B) Sequence alignment of 23 identified *cas12c* and *cas12d* genes; the TnpB-related RuvC catalytic domain location is indicated. C) Plasmid transformation assay showing transformants observed in the presence of different CRISPR-Cas expression constructs or controls including a plasmid-targeting (T) or nontargeting (NT) crRNA.

Supplemental Figure 2 | Cas 12c/d requires a new kind of tracrRNA for DNA interference.

A) Alignment of the conserved segments of Cas12d and Cas12c CRISPR array repeat sequences; universally conserved (black) and partially conserved (gray) nucleotides are indicated; sequences omitted from this analysis, including Cas12d5, were too divergent for high-confidence alignment. B) Partial alignment of Cas12d and Cas12c scoutRNA sequences; universally conserved (black) and partially conserved (gray) nucleotides are indicated; red bar, 5-nt. sequence found in every Cas12d scoutRNA and similar region in Cas12c scoutRNAs. C) Secondary structure predictions (https://rna.urmc.rochester.edu/) for crRNA-tracrRNA sequence-complementary regions for canonical Cas9, Cas12e and Cas12b CRISPR-Cas systems; secondary structure predictions for the base-pairing potential of full-length and truncated sequences of scoutRNA and crRNA from an exemplary Cas12d system.

Supplemental Figure 3 | scoutRNA supports Cas12d-catalyzed and crRNA-targeted DNA cleavage.

A) Size exclusion chromatogram and analytical denaturing polyacrylamide gel analysis of pooled Cas12d protein (CasY.15) fractions obtained from final purification step of protein used in biochemical assays. B) ScoutRNA is essential for Cas12d-mediated on dsDNA cleavage. Cleavage image is shown above, while quantitative plot is shown below. C) Cas12c can not cleave either ssDNA or dsDNA. D) Mapping cleavage site of a 53-base pair DNA substrate. Target sequence is shown as lowercase letters and PAM site is bold, boxed, T=target strand, NT=nontarget strand, T/NT means annealed dsDNA, * indicates 5'-end radiolabeled,and cleavage positions are marked by vertical lines. E) Time-course from the *cis*-cleavages of ssDNA, dsDNA and ssRNA mediated by Cas12d. F) Time-course from the *trans*-cleavages of ssDNA, dsDNA and ssRNA mediated by Cas12d. Cleavage images is show above, while plots of quantified reaction data are shown below.

Supplemental Figure 4 | A short conserved sequence in scoutRNA is required for dsDNA targeting.

A) Time-course of dsDNA cleavages mediated by Cas12d (CasY.15) with crRNA (396) in presence of different scoutRNAs (#1 is wild-type, while #2-5 are different mutations). 396 is wild-type scoutRNA. Cleavage image is shown on top, while different scoutRNAs are shown on bottom. Mutated sequences indicated in pink. B) Filfter binding image used for the plots shown in Figure 4B. 396 is sccoutRNA, 382, 386 and 391are crRNAs with different length. C) Sequence segment sets showing wild-type and reciprocal mutations for Cas12d and SpCas9, used in Fig. 4C.

Supplemental Figure 5 | An RNase III-independent dual RNA-guided pre-crRNA processing mechanism.

A) Time courses of pre-crRNA cleavages mediated by Cas12c. 5'-end radiolabeled 58-nucleotide pre-crRNA substrate is shown in Figure 5; scoutRNA constructs used in each reaction are shown below, with mutated region indicated in pink. B) Cas12d does not process pre-crRNAs under current reaction conditions. ScoutRNA used here is 396; 382, 386, 387, and 391 are different crRNAs and their sequences are given in STAR Methods.

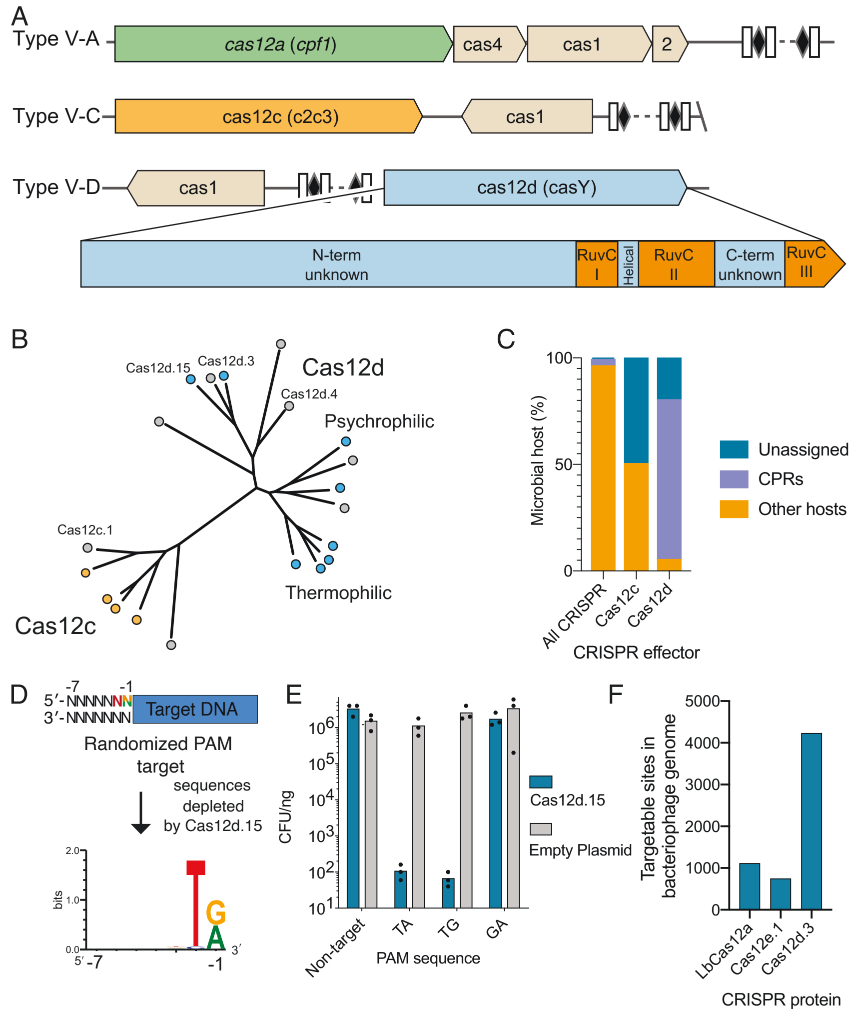

A
Type V-A
cas12a (cpf1)
cas4
cas1
2
Type V-C
cas12c (c2c3)
cas1
Type V-D
cas1
cas12d (casY)
N-term unknown
RuvC I
Helical
RuvC II
C-term unknown
RuvC III
B
Cas12d.15
Cas12d.3
Cas12d
Cas12d.4
Psychrophilic
Cas12c.1
Thermophilic
Cas12c
C
Microbial host (%)
Unassigned
CPRs
Other hosts
All CRISPR
Cas12c
Cas12d
CRISPR effector
D
5′-NNNNNNN
3′-NNNNNNN
Target DNA
Randomized PAM target sequences depleted by Cas12d.15
bits
E
CFU/ng
Cas12d.15
Empty Plasmid
Non-target
TA
TG
GA
PAM sequence
F
Targetable sites in bacteriophage genome
LbCas12a
Cas12e.1
Cas12d.3
CRISPR protein

Figure 1

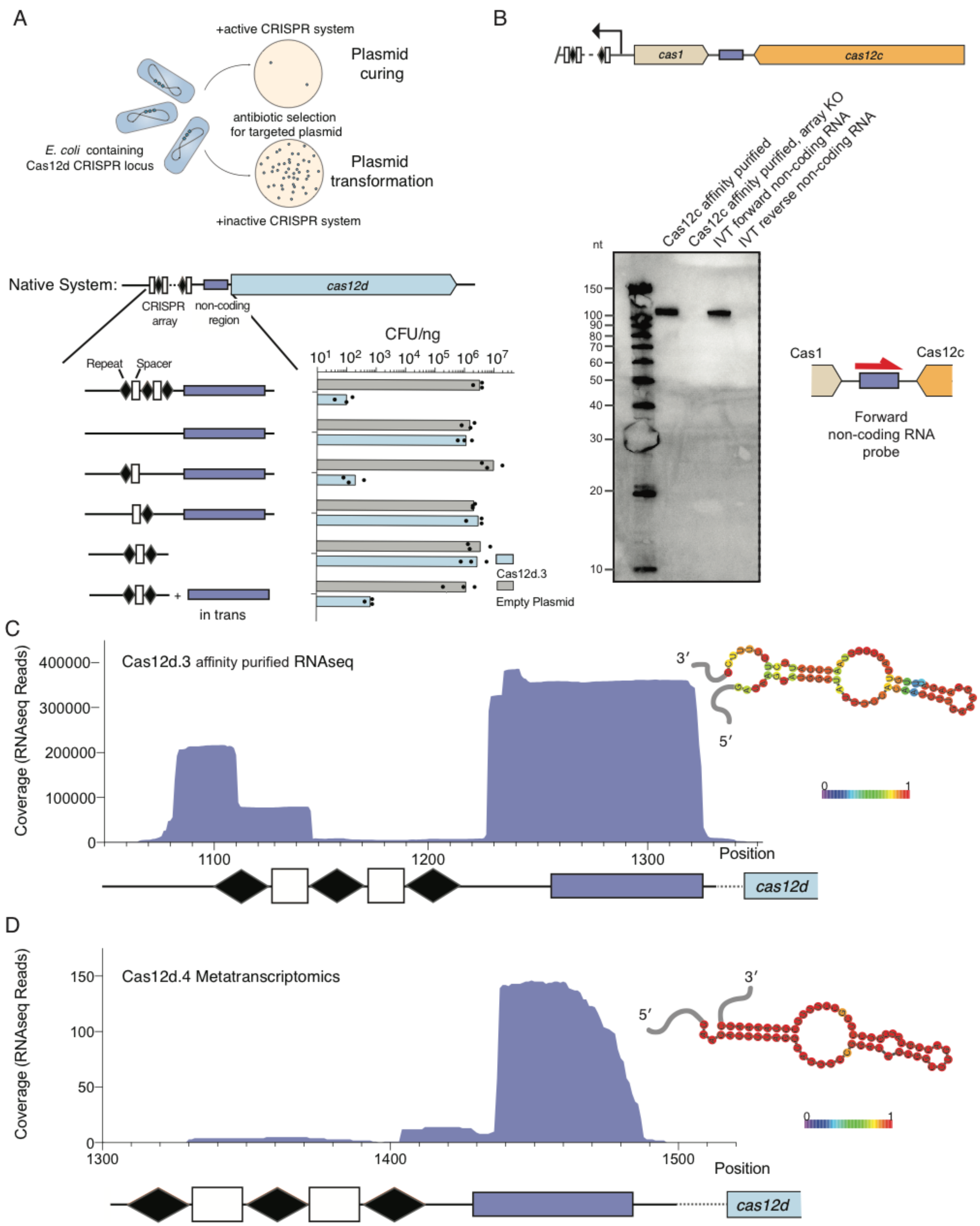

A
+active CRISPR system
Plasmid curing
antibiotic selection for targeted plasmid
E. coli containing Cas12d CRISPR locus
Plasmid transformation
+inactive CRISPR system
Native System:
cas12d
CRISPR array
non-coding region
Repeat
Spacer
in trans
CFU/ng
Cas12d.3
Empty Plasmid
B
cas1
cas12c
Cas12c affinity purified
Cas12c affinity purified, array KO
IVT forward non-coding RNA
IVT reverse non-coding RNA
nt
Cas1
Cas12c
Forward non-coding RNA probe
C
Cas12d.3 affinity purified RNAseq
Coverage (RNAseq Reads)
Position
cas12d
D
Cas12d.4 Metatranscriptomics
Coverage (RNAseq Reads)
Position
cas12d

Figure 2

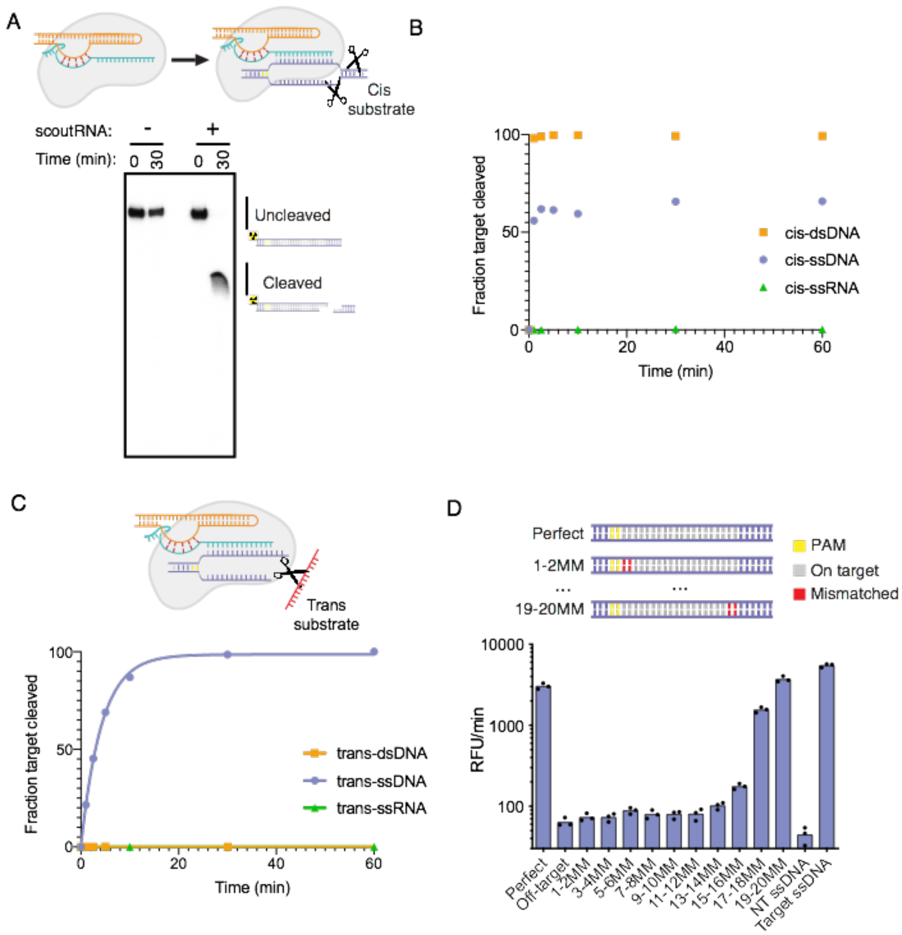
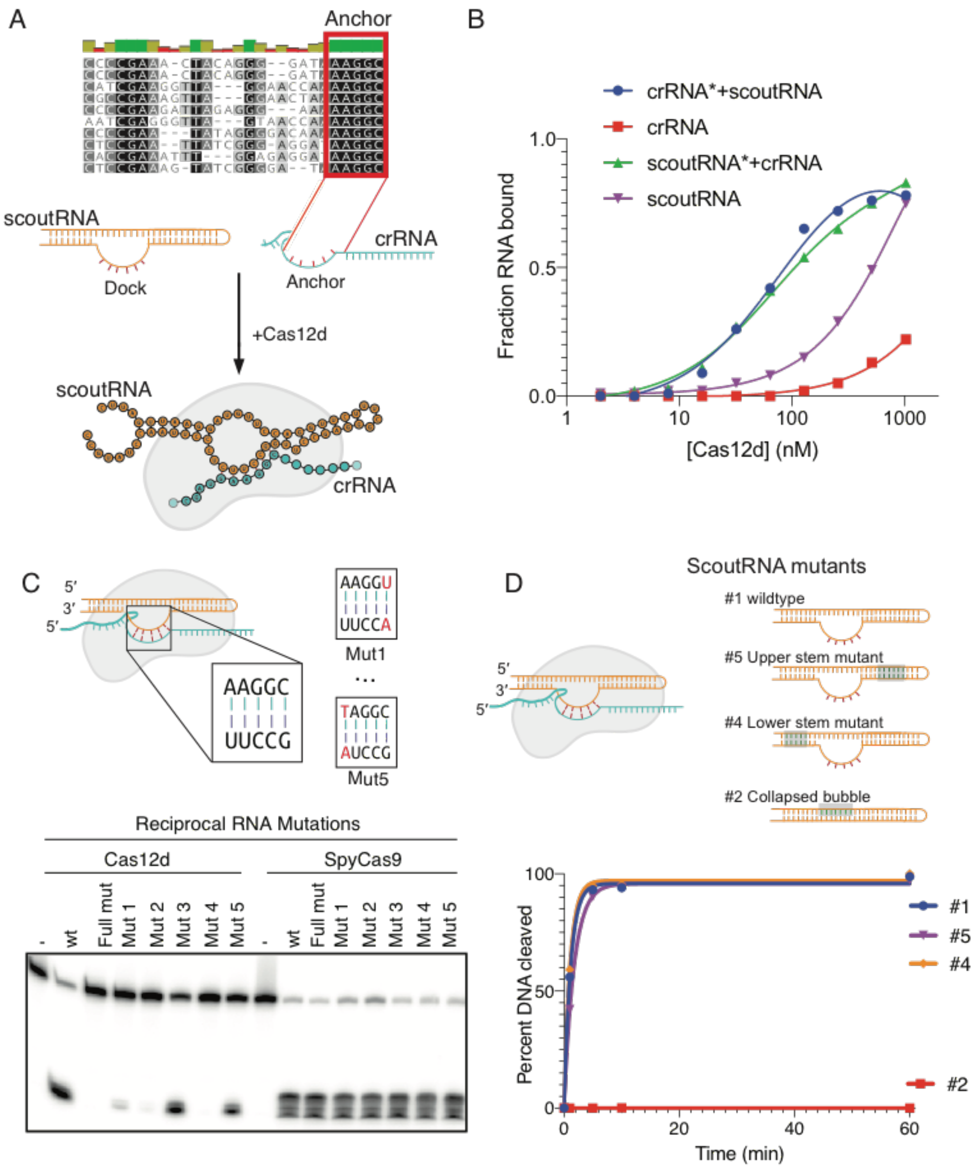
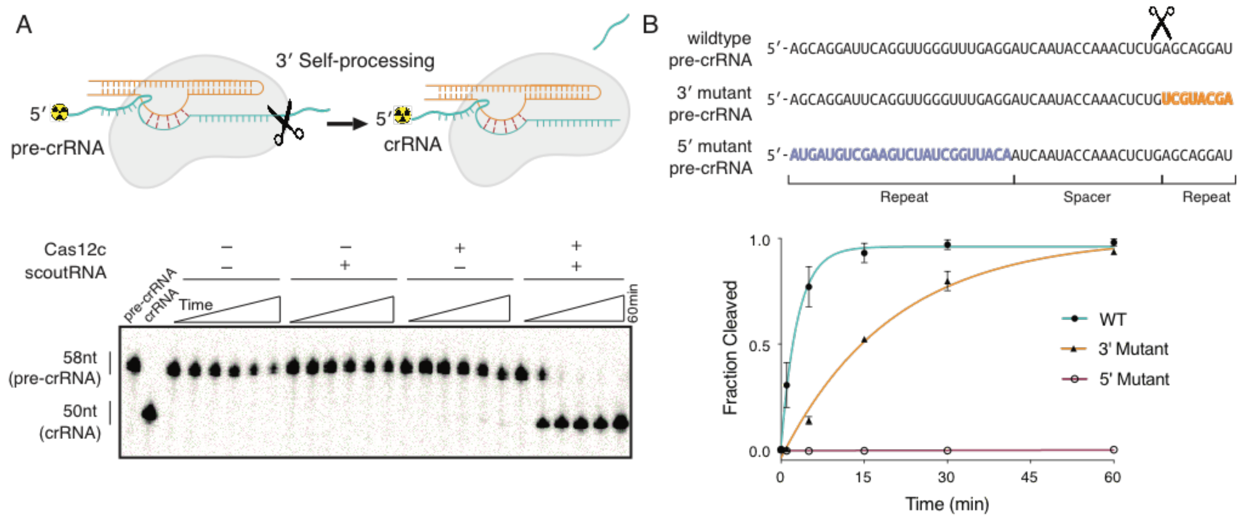

A
Cis
substrate
scoutRNA:
-
+
Time (min):
0
30
0
30
Uncleaved
Cleaved
B
Fraction target cleaved
100
50
0
0
20
40
60
Time (min)
cis-dsDNA
cis-ssDNA
cis-ssRNA
C
Trans
substrate
Fraction target cleaved
100
50
0
0
20
40
60
Time (min)
trans-dsDNA
trans-ssDNA
trans-ssRNA
D
Perfect
1-2MM
...
...
19-20MM
PAM
On target
Mismatched
RFU/min
10000
1000
100
Perfect
Off-target
1-2MM
3-4MM
5-6MM
7-8MM
9-10MM
11-12MM
13-14MM
15-16MM
17-18MM
19-20MM
NT ssDNA
Target ssDNA

Figure 3



A
Anchor
scoutRNA
crRNA
Dock
Anchor
+Cas12d
scoutRNA
crRNA
B
crRNA*+scoutRNA
crRNA
scoutRNA*+crRNA
scoutRNA
Fraction RNA bound
[Cas12d] (nM)
C
AAGGC
UUCCG
AAGGU
UUCCA
Mut1
TAGGC
AUCCG
Mut5
Reciprocal RNA Mutations
Cas12d
SpyCas9
wt
Full mut
Mut 1
Mut 2
Mut 3
Mut 4
Mut 5
D
ScoutRNA mutants
#1 wildtype
#5 Upper stem mutant
#4 Lower stem mutant
#2 Collapsed bubble
Percent DNA cleaved
Time (min)

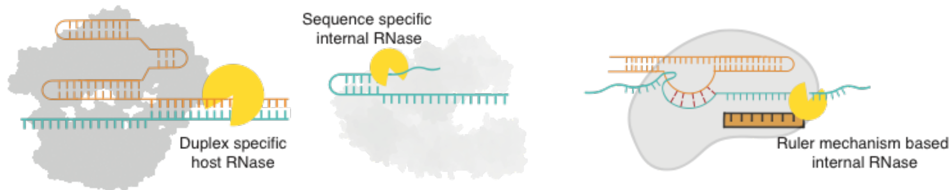
Figure 4



A
3′ Self-processing
5′
pre-crRNA
5′
crRNA
Cas12c
scoutRNA
pre-crRNA
crRNA
Time
60min
58nt
(pre-crRNA)
50nt
(crRNA)
B
wildtype
pre-crRNA
5′-AGCAGGAUUCAGGUUGGGUUUGAGGAUCAAUACCAAACUCUGAGCAGGAU
3′ mutant
pre-crRNA
5′-AGCAGGAUUCAGGUUGGGUUUGAGGAUCAAUACCAAACUCUGUCGUACGA
5′ mutant
pre-crRNA
5′-AUGAUGUCGAAGUCUAUCGGUUACAAUCAAUACCAAACUCUGAGCAGGAU
Repeat
Spacer
Repeat
Fraction Cleaved
Time (min)
WT
3' Mutant
5' Mutant

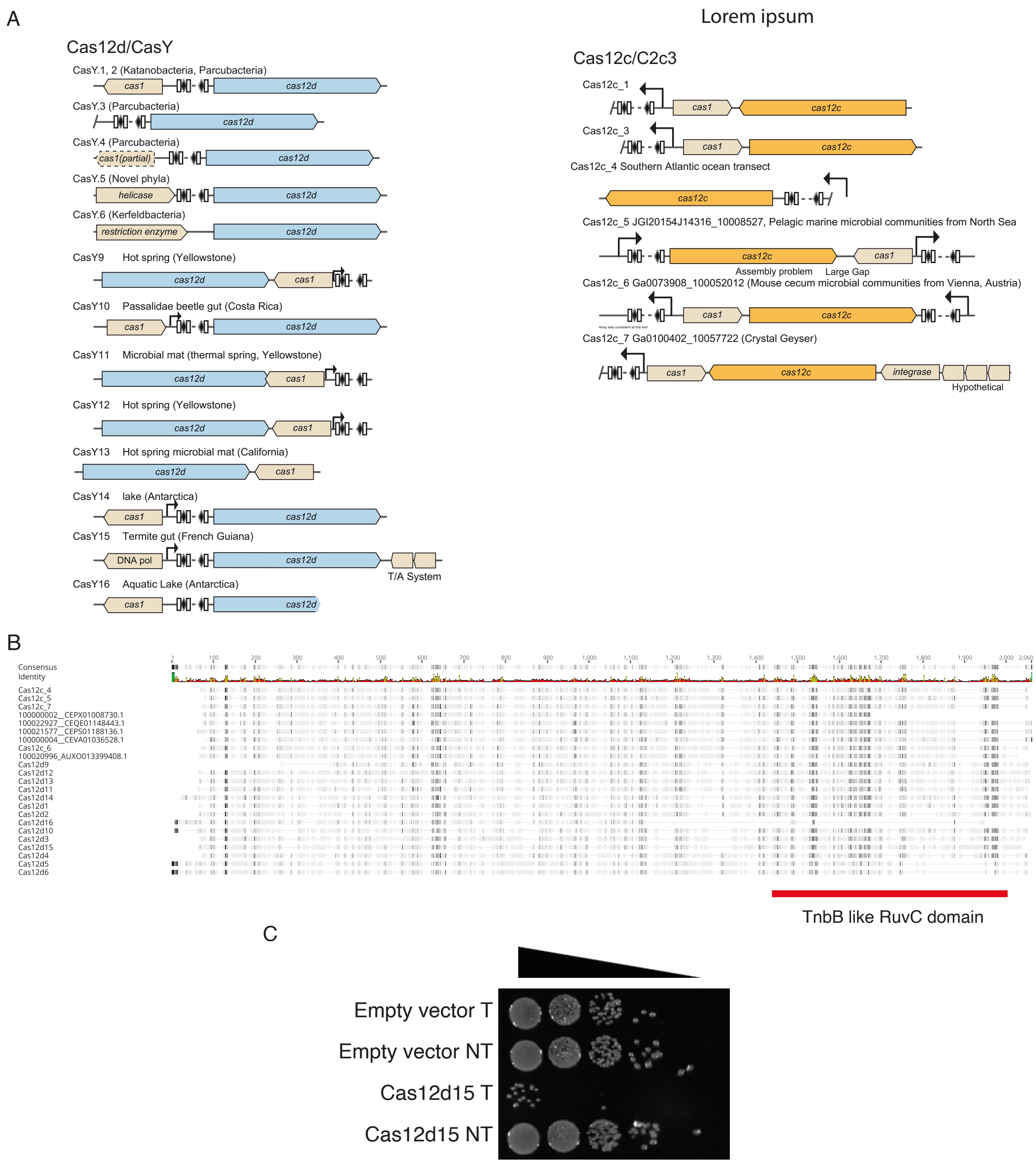
Figure 5

CRISPR Type:	**II (Cas9)**, V-B (Cas12b), V-E (Cas12e)	VI (Cas13), **V-A (Cas12a)**	V-C (Cas12c), V-D (Cas12d)

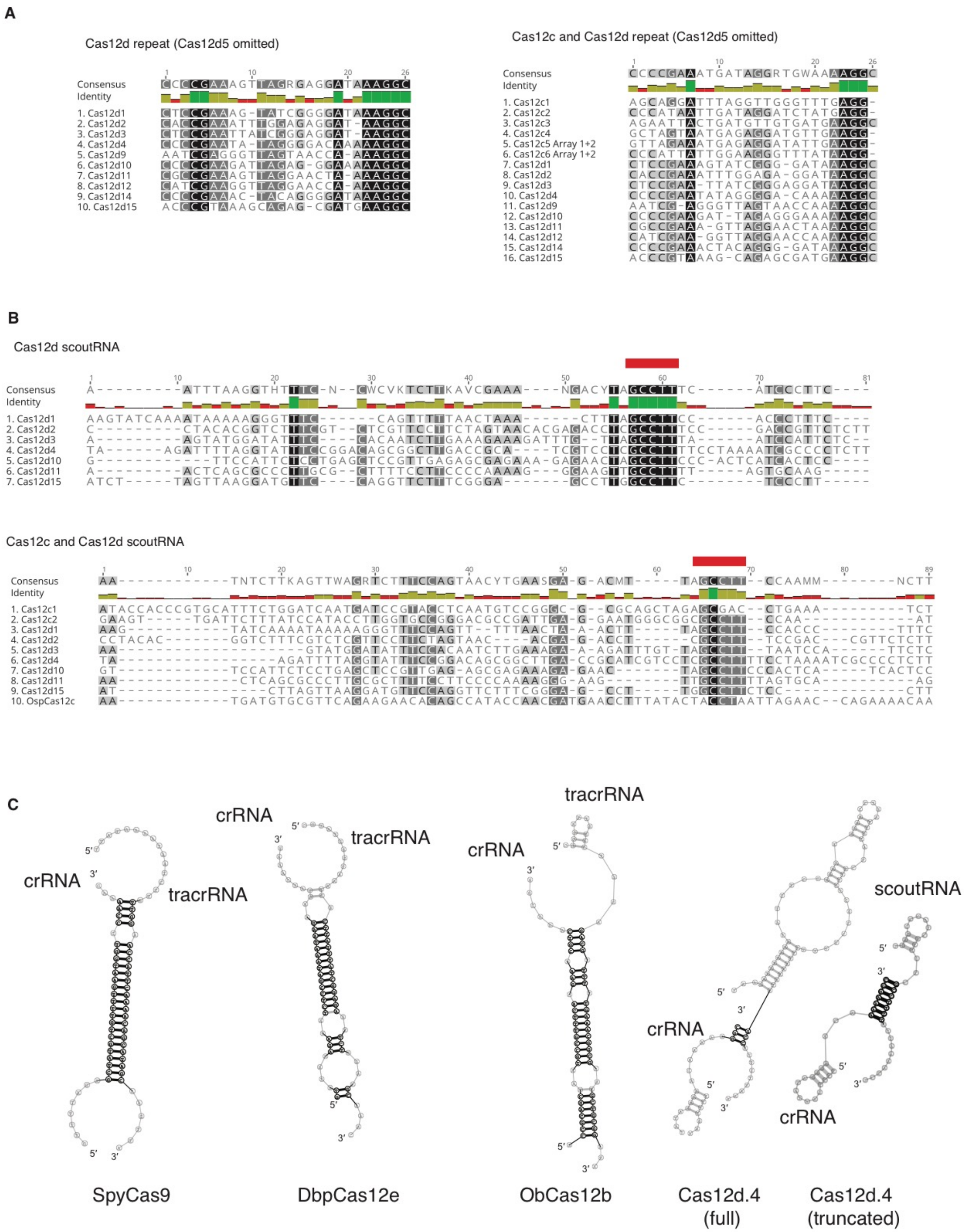


Dual/Single:	Dual	Single	Dual
Processing:	Host RNase III	5′ Self processing	3′ Self processing
Hybridization:	20-40nt	N/A	~5nt
PAM:	Diverse, 3′	T-rich, 5′	TA/TR, 5′

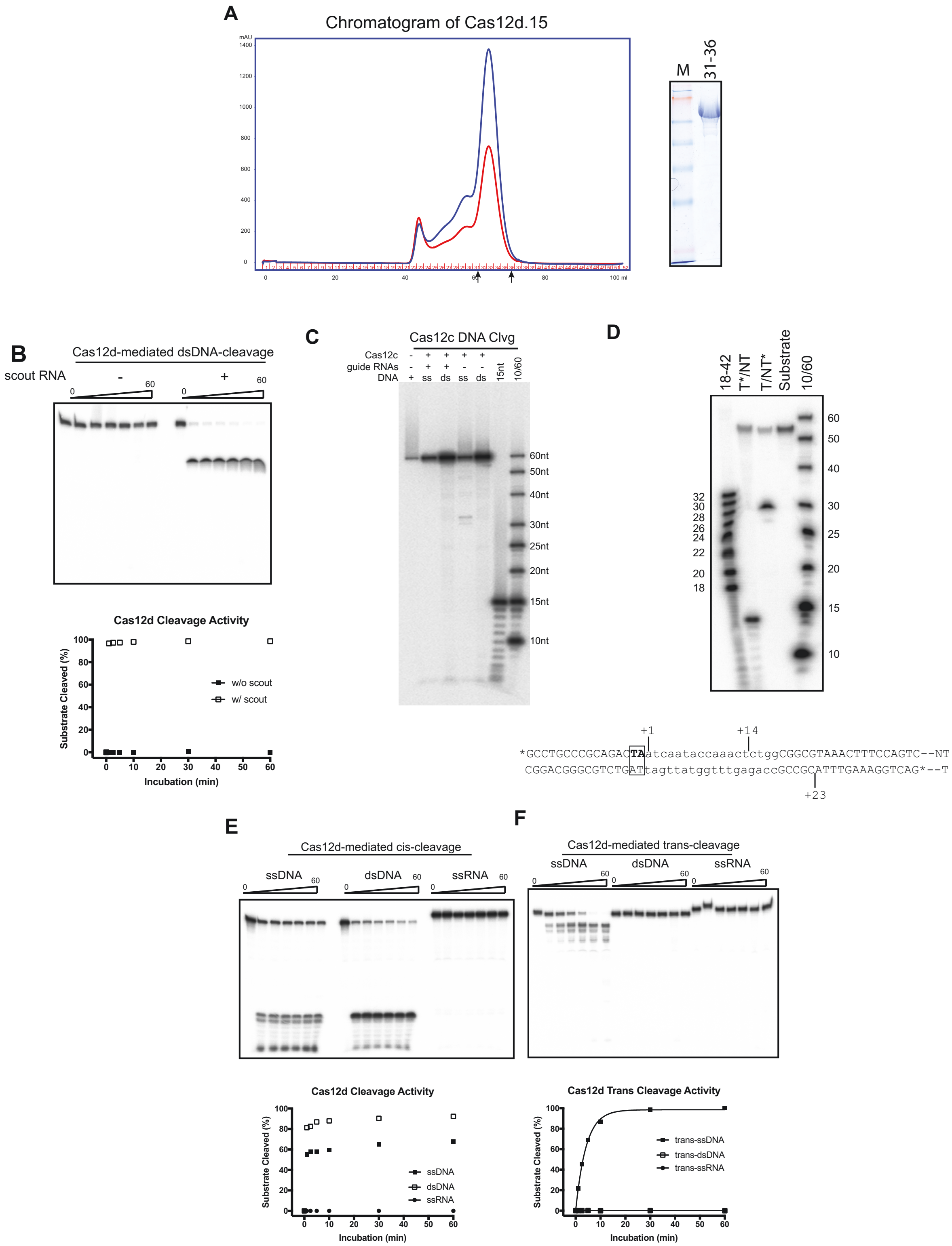
Figure 6



Supplemental Figure 1

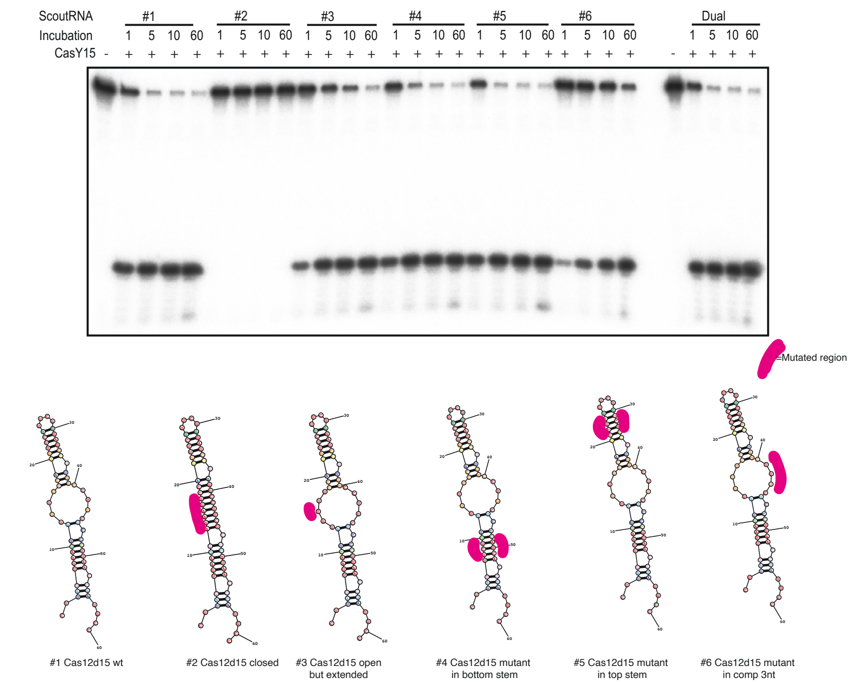


Supplemental Figure 2

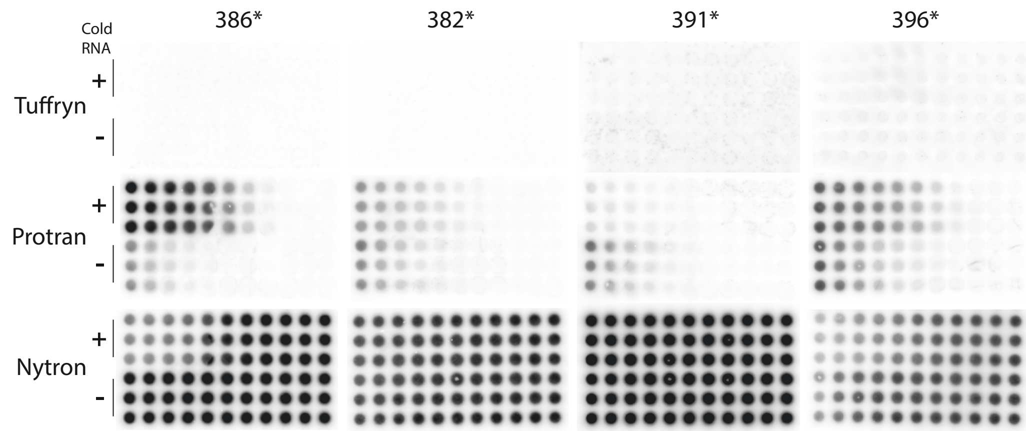


Supplemental Figure 3

A



B



C

Cas12c Reciprocal mutations:

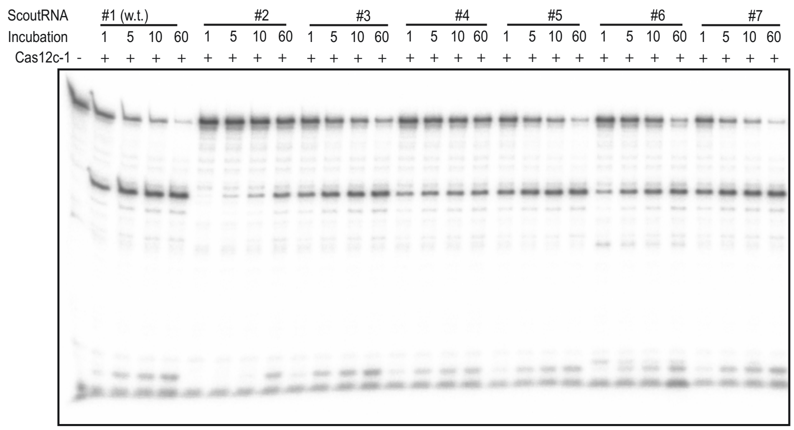
W.T.	Full mut	Mut1	Mut2	Mut3	Mut4	Mut5
GCCUU	ACGCC	ACCUU	GGCUU	GCGUU	GCCAU	GCCUA
CGGAA	UGCGG	UGGAA	CCGAA	CGCAA	CGGUA	CGGAU

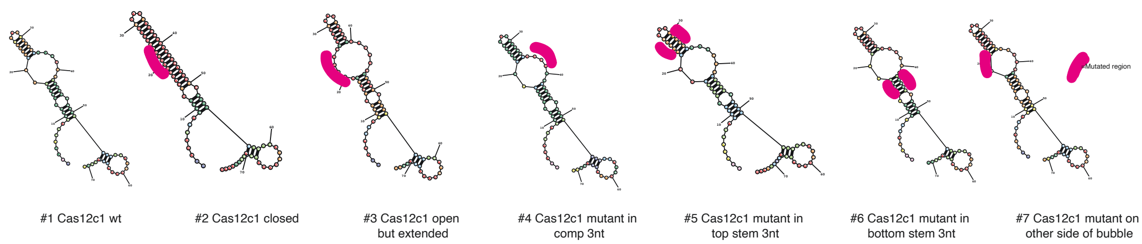
SpCas9 Reciprocal mutations:

W.T.	Full mut	Mut1	Mut2	Mut3	Mut4	Mut5
UUUUA	AAAAU	AUUUA	UAUUA	UUAUA	UUUAA	UUUUU
AAAAU	UUUUA	UAAAU	AUAAU	AAUAU	AAAUU	AAAAA

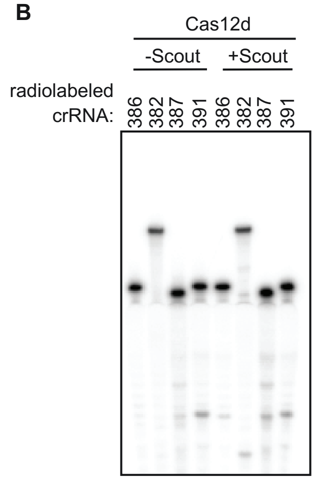
Supplemental Figure 4

A





B



Supplemental Figure 5