PAM recognition by miniature CRISPR-Cas14 triggers programmable double-stranded DNA cleavage

Tautvydas Karvelis[1†*], Greta Bigelyte[1†], Joshua K. Young[2†], Zhenglin Hou[2], Rimante Zedaveinyte[1], Karolina Pociute[1], Arunas Silanskas[1], Česlovas Venclovas[1], Virginijus Siksnys[1*]

Affiliations:

[1]Institute of Biotechnology, Vilnius University, Vilnius, LT-10257, Lithuania.

[2]Corteva Agriscience™ Agriculture Division of DowDuPont™, Johnston, IA 50131, USA.

[†]These authors contributed equally to this work.

[*]Corresponding author. Email: tautvydas.karvelis@bti.vu.lt and siksnys@ibt.lt

Abstract: Small and robust CRISPR-Cas nucleases are highly desirable for genome editing applications. Being guided by a RNA to cleave targets near a short sequence termed a protospacer adjacent motif (PAM), Cas9 and Cas12 offer unprecedented flexibility, however, smaller more compact versions would simplify delivery and extend application. Recently, a new class 2 system encoding a miniature (529 amino acids) effector, Cas14a1, has been shown to exclusively function as a PAM-independent single stranded DNA nuclease. Using biochemical methods, we show that a T-rich PAM sequence triggers Cas14 proteins to also cut double-stranded DNA generating staggered ends. Finally, we demonstrate the ability of Cas14a1 to target and cleave cellular human chromosomal DNA paving the way for genome editing applications with Cas14s.

Main Text: Clustered regularly interspaced short palindromic repeat (CRISPR) associated (Cas) microbial defense systems protect against nucleic acid contagions (1). Utilizing small guide RNAs (gRNAs) transcribed from a CRISPR locus, accessory Cas proteins are directed to silence invading foreign RNA and DNA (2). Based on the number and composition of proteins involved in nucleic acid interference, CRISPR-Cas systems are categorized into distinct classes, 1-2, and types, I-VI (2). Class 2 systems encode a single effector protein for interference and are further subdivided into types, II, V, and VI (3). Cas9 (type II) and Cas12 (type V) proteins have been shown to cleave invading double-stranded (ds) DNA, single-stranded (ss) DNA, and ssRNA (4–9). To recognize and cleave a dsDNA target, both Cas9 and Cas12 require a short sequence, termed the protospacer adjacent motif (PAM), in the vicinity of a DNA sequence targeted by the gRNA (4, 5, 10). Over the past several years, these endonucleases have been adopted as robust genome editing and transcriptome manipulation tools (11, 12). Although both nucleases have been widely used, the size of Cas9 and Cas12 provides constraints on cellular delivery that may limit certain applications including therapeutics (13, 14).

Recently, a new class 2 system with an interference protein nearly half the size of the smallest Cas9 or Cas12 proteins, termed Cas14, has been described (Fig. 1A) (9, 15). Sharing a C-

terminal RuvC domain with Cas12 proteins, Cas14s have been classified as type V effectors (15). Compared to Cas12 orthologs, the N-terminal half of Cas14 differs significantly in length accounting for most of the size difference between the two families (Fig. 1A and Fig. S1). Additionally, Cas14 was shown to differ with respect to the types of nucleic acid substrates it targets. While most Cas12s bind and cut both ssDNA and dsDNA, Cas14 has been reported to exclusively function as a ssDNA nuclease (15). Directed by its gRNA, it was shown to bind and cleave a ssDNA target in a PAM-independent manner. Additionally, target binding and cleavage by Cas14 triggered collateral nuclease activity that manifested as trans-acting non-specific ssDNA degradation (15). This seems to be a feature largely shared by the type V family although Cas12g was reported to initially target ssRNA and then indiscriminately degrade both ssDNA and ssRNA (8, 9).

Here, we report that Cas14, like Cas9 and most type V (a-e, h, i) proteins, cleaves dsDNA targets in the vicinity of a T-rich PAM. Tests were performed using the intact Cas14 CRISPR-Cas system from *Candidatus Micrarchaeota archaeon* (CG1_02_47_40) (GeneBank: MK005740.1), termed Cas14b5 (15). First, spacers capable of targeting a PAM plasmid library (16) were incorporated into its CRISPR array. The resulting modified CRISPR-Cas14b5 locus was then synthesized, cloned into a low copy plasmid, and transformed into *E. coli*. Next, a previously developed PAM determination assay (16, 17) was adapted to assess the ability of the Cas14b5 to recognize and cleave a dsDNA target (Fig. 1B). This was accomplished by combining clarified lysate containing Cas14b5 protein and guide RNAs expressed from the reengineered locus with the PAM library. Then DNA breaks were captured by double-stranded adapter ligation, enriched for by PCR, and Illumina deep sequenced as described previously (16, 17). DNA cleavage occurring in the target sequence was evaluated by scanning regions in the protospacer for elevated frequencies of adapter ligation relative to negative controls (experiments assembled with lysate from *E. coli* not transformed with the Cas14b5 locus). As illustrated in Fig. S2A, adapters were more likely to be ligated after the 21st protospacer position 3' of the randomized PAM. Next, fragments resulting from DNA cleavage in the target region, were evaluated for PAM preferences. Assuming 5'PAM recognition, a T-rich sequence (5'-TTAT-3') was recovered in the vicinity of the gRNA target only in the Cas14b5 treated sample (Fig. 1C and Fig. S2B). To confirm the PAM sequence and dsDNA cleavage position, a plasmid was constructed containing a target adjacent to the identified 5'-TTAT-3' PAM sequence and subjected to biochemical cleavage experiments using cell lysate. Also, to increase Cas14b5 concentration, a higher copy number DNA expression plasmid equipped with an inducible T7 promoter was utilized. Sequencing of the target plasmid cleavage products confirmed cleavage at the 21st position of the target sequence in respect to the PAM, especially, for reactions where expression of the Cas14b5 locus was regulated by T7 (Fig. S2C). Reactions using deletion variants further confirmed that Cas14b5 was the sole protein required for the observed dsDNA target recognition and cleavage activity (Fig. S2C).

To corroborate our findings, the model Cas14a1 system was chosen (15). Like Cas14b5 experimentation, *E. coli* lysate from cells expressing Cas14a1 was used. However, instead of

expressing gRNAs from a modified CRISPR-Cas locus, an *in vitro* transcribed single guide RNA (sgRNA) capable of targeting the PAM library was included in the reaction mixture. Next, cleavage products were captured and analyzed as described above (Fig. 1B). Akin to Cas14b5, a cleavage signal just after the 24th protospacer position was detected (Fig. S3A) and a T-rich PAM sequence, 5'-TTTR-3', immediately 5' of the sgRNA target was recovered (Fig. 1D and Fig. S3B). Taken together, this shows that like Cas14b5, Cas14a1 also recognizes and cuts dsDNA in a PAM-dependent manner.

Next, programmable Cas14a1 dsDNA cleavage was studied using purified components. First, to ensure favorable reaction conditions, the effect of sgRNA spacer length, temperature, salt concentration, and divalent metal ions were evaluated on Cas14a1 dsDNA cleavage activity (Fig. S4). Experiments revealed that Cas14a1 ribonucleoprotein (RNP) complex is a Mg^{2+}-dependent endonuclease that functions best in low salt concentrations (5-25 mM) and in a temperature range of ~46-50°C (Fig. S4). Furthermore, sgRNA spacers of around 20 nt supported the most robust dsDNA cleavage activity (Fig. S4). Under optimized reaction conditions, supercoiled (SC) plasmid DNA containing a target sequence flanked by a Cas14a1 PAM, 5'-TTTA-3' was completely converted to a linear form (FLL), thus, illustrating the formation of a dsDNA break (Fig. 2A). Additionally, cleavage of linear DNA resulted in DNA fragments of expected size further validating Cas14a1 mediated dsDNA break formation. Next, we confirmed that Cas14a1 requires both PAM and sgRNA recognition to cleave a dsDNA target (Fig. S5). Finally, alanine substitution of conserved RuvC active site residues abolished cutting activity further confirming that the RuvC domain is essential for the observed dsDNA cleavage activity (Fig. 2B).

The type of dsDNA break generated by Cas14a1 was examined next. Using run-off sequencing, we observed that Cas14a1 makes 5' staggered overhanging DNA cut-sites. Cleavage predominantly occurred centered around positions 20-24 bp in respect to PAM-sequence (Fig. 2C) independent of spacer length and plasmid topology (Fig. S6). Next, the cleavage pattern of Cas14a1 was assessed on synthetic double-stranded oligodeoxynucleotides. As illustrated in Fig. 2D and Fig. S7, a 5' staggered cut pattern albeit with less strictly defined cleavage positions than observed with larger DNA fragments was also noted.

Next, we investigated if the non-specific ssDNA degradation activity of Cas14a1 could be induced not just by ssDNA targets (15) but also by dsDNA targets. First, the ability of Cas14a1 to indiscriminately degrade single-stranded M13 DNA in the presence of a ssDNA target without a PAM was confirmed. Next, a dsDNA target containing a 5' PAM and sgRNA target for Cas14a1 was also tested for its ability to trigger non-selective ssDNA degradation. As shown in Fig. 2E, the trans-acting ssDNase activity of Cas14a1 was activated by both ssDNA and dsDNA targets, analogous to observations made for Cas12a (8).

Finally, the ability of Cas14a1 to target and cleave human chromosomal dsDNA in HEK293 cells was examined. First, NLS-tagged Cas14a1 and sgRNA ribonucleoprotein (RNP) complexes capable of targeting sites adjacent to a 5' TTTR PAM in the VEGFA2 or 3 loci were prepared. Next, they were delivered directly into HEK293 cells using cationic lipid-based transfection.

After 72 hours, VEGFA2 and 3 target sites were assayed by targeted deep sequencing for the presence of insertion or deletion (indel) mutations indicative of dsDNA cleavage and cellular repair. For all Cas14a1-sgRNA replicates, indel mutations were recovered within the expected cleavage window for Cas14a1 (Fig. 3A and B) and the results were significant relative to the negative controls (lipofections where the sgRNA was omitted) (Fig. S8).

Taking into consideration size similarities between Cas14 and type V ancestral TnpB-like transposon associated proteins (3), it was speculated that Cas14 hadn't evolved the necessary structures to facilitate RNA-guided dsDNA target recognition and cleavage (15). Here, using biochemical approaches, we provided evidence, altogether, demonstrating that Cas14 is a programmable nuclease capable of introducing targeted dsDNA breaks similar to much larger Cas12 proteins. First, we uncovered T-rich PAM recognition that supports dsDNA cleavage similar to other type V interference proteins (9, 10, 18) for two Cas14 family members, Cas14a1 and Cas14b5. Additionally, using purified protein and sgRNA, we reconstituted dsDNA target cleavage for Cas14a1 and confirmed that both 5' PAM and sgRNA recognition are required for dsDNA cleavage. Despite the small size of Cas14, we also showed PAM-distal dsDNA staggered cleavage for Cas14a1 reminiscent of larger Cas12 proteins (10, 19–22). Moreover, as described for Cas12a, we identified that both ssDNA as well as PAM containing dsDNA targets can serve as triggers for Cas14 non-specific ssDNase activity, potentially expanding the utility of recently developed nucleic acid detection platforms like SHERLOCK and DETECTR (8, 23, 24). Finally, we illustrate that Cas14 is capable of cleaving chromosomal DNA in eukaryotic cells suggesting that it can be harnessed as a genome editing tool.

References:

1. Koonin,E. V., Makarova,K.S. and Wolf,Y.I. (2017) Evolutionary Genomics of Defense Systems in Archaea and Bacteria. *Annu. Rev. Microbiol.*, **71**, 233–261.
2. Koonin,E. V., Makarova,K.S. and Zhang,F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
3. Shmakov,S., Smargon,A., Scott,D., Cox,D., Pyzocha,N., Yan,W., Abudayyeh,O.O., Gootenberg,J.S., Makarova,K.S., Wolf,Y.I., *et al.* (2017) Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **15**, 169–182.
4. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci.*, **109**, 2579–2586.
5. 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.
6. Ma,E., Harrington,L.B., O'Connell,M.R., Zhou,K. and Doudna,J.A. (2015) Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Mol. Cell*, **60**, 398–407.
7. Zhang,Y., Rajan,R., Seifert,H.S., Mondragón,A. and Sontheimer,E.J. (2015) DNase H Activity of Neisseria meningitidis Cas9. *Mol. Cell*, **60**, 242–255.
8. 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.
9. 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.
10. 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.
11. Knott,G.J. and Doudna,J.A. (2018) CRISPR-Cas guides the future of genetic engineering. *Science*, **361**, 866–869.
12. Klompe,S.E. and Sternberg,S.H. (2018) Harnessing "A Billion Years of Experimentation": The Ongoing Exploration and Exploitation of CRISPR–Cas Immune Systems. *Cris. J.*, **1**, 141–158.
13. Wu,Z., Yang,H. and Colosi,P. (2010) Effect of genome size on AAV vector packaging. *Mol. Ther.*, **18**, 80–6.
14. Lino,C.A., Harper,J.C., Carney,J.P. and Timlin,J.A. (2018) Delivering CRISPR: a review of the challenges and approaches. *Drug Deliv.*, **25**, 1234–1257.
15. Harrington,L.B., Burstein,D., Chen,J.S., Paez-Espino,D., Ma,E., Witte,I.P., Cofsky,J.C., Kyrpides,N.C., Banfield,J.F. and Doudna,J.A. (2018) Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science*, **362**, 839–842.

16. Karvelis,T., Gasiunas,G., Young,J., Bigelyte,G., Silanskas,A., Cigan,M. and Siksnys,V. (2015) Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.*, **16**, 253.

17. Karvelis,T., Young,J.K. and Siksnys,V. (2019) A pipeline for characterization of novel Cas9 orthologs. In *Methods in Enzymology*. Elsevier Inc., Vol. 616, pp. 219–240.

18. Liu,J.-J., Orlova,N., Oakes,B.L., Ma,E., Spinner,H.B., Baney,K.L.M., Chuck,J., Tan,D., Knott,G.J., Harrington,L.B., *et al.* (2019) CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*, **566**, 218–223.

19. Swarts,D.C., van der Oost,J. and Jinek,M. (2017) Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell*, **66**, 221-233.e4.

20. Strohkendl,I., Saifuddin,F.A., Rybarski,J.R., Finkelstein,I.J. and Russell,R. (2018) Kinetic Basis for DNA Target Specificity of CRISPR-Cas12a. *Mol. Cell*, **71**, 816-824.e3.

21. 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.

22. Stella,S., Alcón,P. and Montoya,G. (2017) Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature*, **546**, 559–563.

23. Gootenberg,J.S., Abudayyeh,O.O., Lee,J.W., Essletzbichler,P., Dy,A.J., Joung,J., Verdine,V., Donghia,N., Daringer,N.M., Freije,C.A., *et al.* (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science*, **356**, 438–442.

24. Gootenberg,J.S., Abudayyeh,O.O., Kellner,M.J., Joung,J., Collins,J.J. and Zhang,F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, **360**, 439–444.

25. Stormo,G.D. (2013) Modeling the specificity of protein-DNA interactions. *Quant. Biol.*, **1**, 115–130.

Acknowledgments: We would like to thank S. Dooley, S. Paulraj, H. Lin, and M. King for informatics support and S. Gasior, V. Djukanovic, C. Skrdlant, and G. Zastrow-Hayes for Illumina sequencing support.

Funding: This work was made possible by a joint collaboration between Vilnius University and Corteva Agriscience.

Author contributions: T.K., J.K.Y. and V.S. conceived and designed the study. T.K., G.B., J.K.Y., R.Z., K.P., A.S. and C.V. designed and performed the experiments. T.K., J.K.Y. and V.S. wrote the manuscript with the input from all authors. All authors read, edited and approved the manuscript.

Competing interests: T.K., J.K.Y., Z.H., and V.S. have filed patent applications related to the manuscript. J.K.Y., Z.H., and H.L. are employees of Corteva Agriscience. V.S. is a Chairman of and has financial interest in CasZyme.

Data and materials availability: All data are available in the manuscript or the supplementary material.

Supplementary Materials:

Materials and Methods

Figures S1-S8

Table S1

Data S1

References

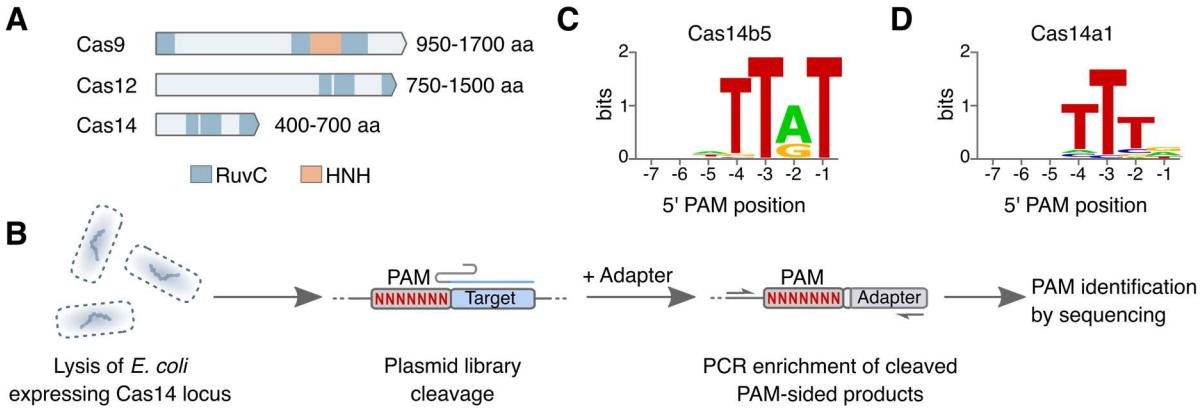

Fig. 1. Detection of dsDNA target recognition and cleavage activity for Cas14 proteins. (**A**) Schematic representation of domain architecture for signature class 2 effector nucleases. (**B**) Workflow of the biochemical approach used to examine PAM recognition and detect dsDNA cleavage. WebLogos of enriched PAM sequences for Cas14b5 (**C**) and Cas14a1 (**D**).

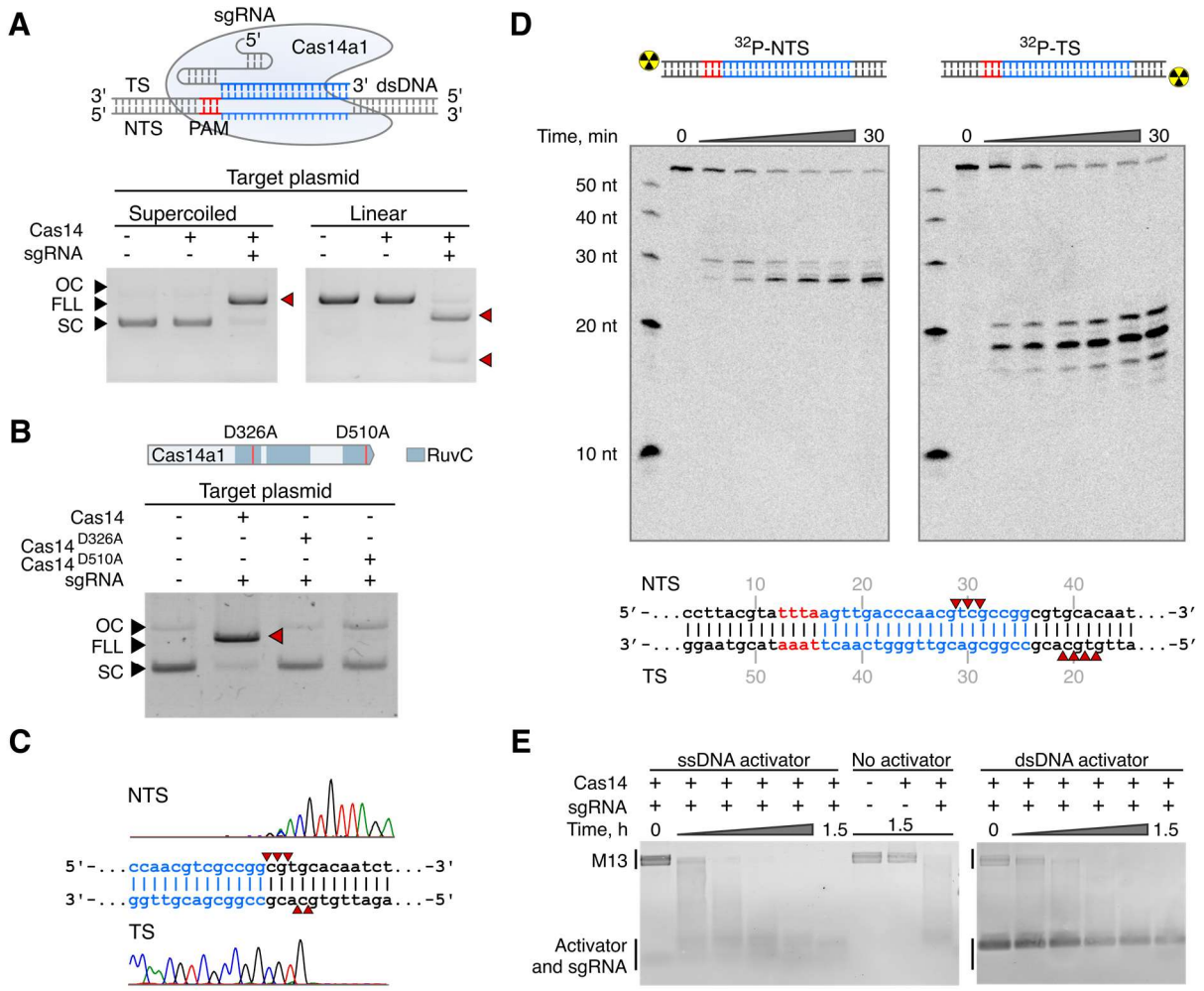

Fig. 2. Cas14a1 RNP complex is a PAM-dependent dsDNA endonuclease. (**A**) Cas14a1 RNP complex cleaves plasmid DNA targets *in vitro* in a PAM-dependent manner. (**B**) Alanine substitution of two conserved RuvC active site residues completely abolishes Cas14a1 DNA cleavage activity. (**C**) Run-off sequencing of the Cas14a1 pre-cleaved plasmid DNA indicates that cleavage is centered around positions 20-24 bp 3' of the PAM resulting in 5'-overhangs. (**D**) Oligoduplex cleavage patterns are consistent with staggered cleavage but differ from experiments assembled with plasmid DNA (shown in C). (**E**) Collateral non-specific M13 ssDNA degradation activity by Cas14a1 triggered by ssDNA and PAM-containing dsDNA. TS – target strand, NTS – non-target strand, SC – supercoiled, FLL – full length linear, OC – open circular.

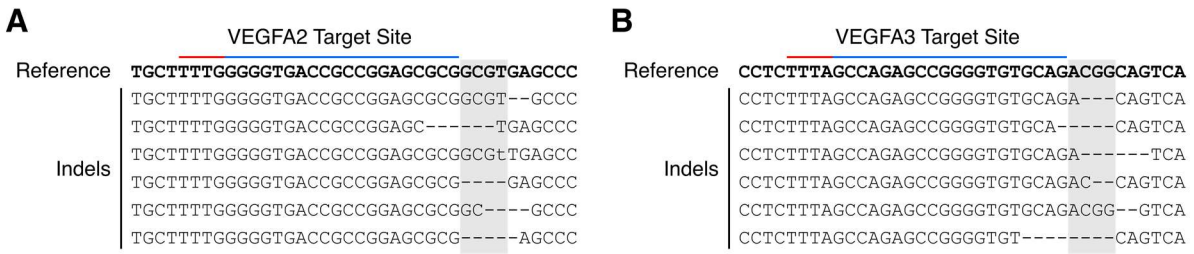

Fig. 3. ***In vivo*** **genome editing using Cas14a1.** Indel mutations recovered in VEGFA2 (**A**) and VEGFA3 (**B**) target sites in HEK293 cells 72 h post Cas14a1 RNP treatment. Dashes represent deletions and gray shaded box indicates expected cleavage window for Cas14a1.