Sciencexpress Report

RNA-Guided Human Genome Engineering via Cas9

Prashant Mali,[1,5] Luhan Yang,[1,3,5] Kevin M. Esvelt,[2] John Aach,[1] Marc Guell,[1] James E. DiCarlo,[4] Julie E. Norville,[1] George M. Church[1,2*]

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA. [2]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02138, USA. [3]Biological and Biomedical Sciences Program, Harvard Medical School, Boston, MA 02115, USA.. [4]Department of Biomedical Engineering, Boston University, Boston, MA 02215, USA.

[5]These authors contributed equally to this work.

*To whom correspondence should be addressed. E-mail: gchurch@genetics.med.harvard.edu

Bacteria and archaea have evolved adaptive immune defenses termed clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) systems that use short RNA to direct degradation of foreign nucleic acids. Here, we engineer the type II bacterial CRISPR system to function with custom guide RNA (gRNA) in human cells. For the endogenous AAVS1 locus, we obtained targeting rates of 10 to 25% in 293T cells, 13 to 8% in K562 cells, and 2 to 4% in induced pluripotent stem cells. We show this process relies on CRISPR components, is sequence-specific, and upon simultaneous introduction of multiple gRNAs, can effect multiplex editing of target loci. We also compute a genome-wide resource of ~190k unique gRNAs targeting ~40.5% of human exons. Our results establish an RNA-guided editing tool for facile, robust, and multiplexable human genome engineering.

Bacterial and archaeal CRISPR systems rely on crRNAs in complex with Cas proteins to direct degradation of complementary sequences present within invading viral and plasmid DNA (*1*–*3*). A recent in vitro reconstitution of the *S. pyogenes* type II CRISPR system demonstrated that crRNA fused to a normally trans-encoded tracrRNA is sufficient to direct Cas9 protein to sequence-specifically cleave target DNA sequences matching the crRNA (*4*). The fully defined nature of this 2-component system suggested that it might function in the cells of eukaryotic organisms such as yeast, plants, and even mammals. By cleaving genomic sequences targeted by RNA sequences (*4*–*6*), such a system could greatly enhance the ease of genome engineering.

Here we engineer the protein and RNA components of this bacterial type II CRISPR system in human cells. We began by synthesizing a human codon-optimized version of the Cas9 protein bearing a C terminus SV40 nuclear localization signal and cloning it into a mammalian expression system (Fig. 1A and fig. S1A). To direct Cas9 to cleave sequences of interest, we expressed crRNA-tracrRNA fusion transcripts, hereafter referred to as guide RNAs (gRNAs), from the human U6 polymerase III promoter. Importantly, directly transcribing gRNAs allowed us to avoid reconstituting the RNA processing machinery employed by bacterial CRISPR systems (Fig. 1A and fig. S1B) (*4*, *7*–*9*). Constrained only by U6 transcription initiating with G and the requirement for the PAM (protospacer-adjacent motif) sequence -NGG following the 20 bp crRNA target, our highly versatile approach can in principle target any genomic site of the form $GN_{20}GG$ (fig. S1C; see SI for a detailed discussion).

To test the functionality of our implementation for genome engineering, we developed a GFP reporter assay (Fig. 1B) in 293T cells similar to one previously described (*10*). Specifically, we established a stable cell line bearing a genomically integrated GFP coding sequence disrupted by the insertion of a stop codon and a 68bp genomic fragment from the AAVS1 locus that renders the expressed protein fragment non-fluorescent. Homologous recombination (HR) using an appropriate repair donor can restore the normal GFP sequence, enabling us to quantify the resulting GFP^+ cells by flow activated cell sorting (FACS).

To test the efficiency of our system at stimulating HR, we constructed two gRNAs, T1 and T2, that target the intervening AAVS1 fragment (Fig. 1B) and compared their activity to that of a previously described TAL effector nuclease heterodimer (TALEN) targeting the same region (*11*). We observed successful HR events using all three targeting reagents, with gene correction rates using the T1 and T2 gRNAs approaching 3% and 8% respectively (Fig. 1C). This RNA-mediated editing process was notably rapid, with the first detectable GFP^+ cells appearing ~20 hours post transfection compared to ~40 hours for the AAVS1 TALENs. We observed HR only upon simultaneous introduction of the repair donor, Cas9 protein, and gRNA, confirming that all components are required for genome editing (fig. S2). While we noted no apparent toxicity associated with Cas9/crRNA expression, work with ZFNs and TALENs has shown that nicking only one strand further reduces toxicity. Accordingly, we also tested a Cas9D10A mutant that known to function as a nickase in vitro, which yielded similar HR but lower non-homologous end joining (NHEJ) rates (fig. S3) (*4*, *5*). Consistent with (*4*) where a related Cas9 protein is shown to cut both strands 6 bp upstream of the PAM, our NHEJ data confirmed that most deletions or insertions occurred at the 3′ end of the target sequence (fig. S3B). We also confirmed that mutating the target genomic site prevents the gRNA from effecting HR at that locus, demonstrating that CRISPR-mediated genome editing is sequence specific (fig. S4). Finally, we showed that two gRNAs targeting sites in the GFP gene, and also three additional gRNAs targeting fragments from homologous regions of the DNA methyl transferase 3a (DNMT3a) and DNMT3b genes could sequence specifically induce significant HR in the engineered reporter cell lines (figs. S5 and S6). Together these results confirm that RNA-guided genome targeting in human cells is simple to execute and induces robust HR across multiple target sites.

Having successfully targeted an integrated reporter, we next turned to modifying a native locus. We used the gRNAs described above to target the AAVS1 locus located in the PPP1R12C gene on chromosome 19, which is ubiquitously expressed across most tissues (Fig. 2A) in 293Ts, K562s, and PGP1 human iPS cells (*12*) and analyzed the results by next-generation sequencing of the targeted locus. Consistent with our results for the GFP reporter assay, we observed high numbers of NHEJ events at the endogenous locus for all three cell types. The two gRNAs T1 and T2 achieved NHEJ rates of 10 and 25% in 293Ts, 13 and 38% in K562s, and 2 and 4% in PGP1-iPS cells, respectively (Fig. 2B). We observed no overt toxicity from the Cas9 and crRNA expression required to induce NHEJ in any of these cell types (fig. S7). As expected, NHEJ-mediated deletions for T1 and T2 were centered around the target site positions, further validating the sequence specificity of this targeting process (figs. S7 to S9). Interestingly, simultaneous introduction of both T1 and T2 gRNAs resulted in high efficiency deletion of the intervening

19bp fragment (fig. S8), demonstrating that multiplexed editing of genomic loci is feasible using this approach.

Lastly, we attempted to use HR to integrate either a dsDNA donor construct (*13*) or an oligo donor into the native AAVS1 locus (Fig. 2C and fig. S10). We confirmed HR-mediated integration using both approaches by PCR (Fig. 2D and fig. S10) and Sanger sequencing (Fig. 2E). We also readily derived 293T or iPS clones from the pool of modified cells using puromycin selection over two weeks (Fig. 2F and fig. S10). These results demonstrate that Cas9 is capable of efficiently integrating foreign DNA at endogenous loci in human cells.

Our versatile RNA-guided genome editing system can be readily adapted to modify other genomic sites by simply modifying the sequence of our gRNA expression vector to match a compatible sequence in the locus of interest. To facilitate this process, we bioinformatically generated ~190,000 specifically gRNA-targetable sequences targeting ~40.5% exons of genes in the human genome (refer to Methods and table S1). We incorporated these target sequences into a 200bp format compatible with multiplex synthesis on DNA arrays (*14*) (fig. S11 and tables S2 and S3). This resource provides a ready genome-wide reference of potential target sites in the human genome and a methodology for multiplex gRNA synthesis.

Our results demonstrate the promise of CRISPR-mediated gene targeting for RNA guided, robust and multiplexable mammalian genome engineering. The ease of retargeting our system to modify genomic sequences greatly exceeds that of comparable zinc finger nucleases (ZFNs) and TALENs while offering similar or greater efficiencies (*4*). Existing studies of Type II CRISPR specificity (*4*) suggest that target sites must perfectly match the PAM sequence NGG and the 8-12 base "seed sequence" at the 3′ end of the gRNA. The importance of the remaining 8-12 bases is less well understood and may depend on the binding strength of the matching gRNAs or on the inherent tolerance of Cas9 itself. Indeed, Cas9 will tolerate single mismatches at the 5′ end in bacteria and in vitro, suggesting that the 5′ G is not required. Moreover, it is likely the target locus's underlying chromatin structure and epigenetic state will also impact the efficiency of genome editing in eukaryotic cells (*13*), although we suspect that Cas9's helicase activity may render it more robust to these factors: but this remains to be evaluated. Elucidating the frequency and underlying causes of off-target nuclease activity (*15*, *16*) induced by CRISPR, ZFN (*17*, *18*) and TALEN (*19*, *20*) genome engineering tools will be of utmost importance for safe genome modification and perhaps gene therapy. Potential avenues for improving CRISPR specificity include evaluating Cas9 homologs identified through bioinformatics and directed evolution of these nucleases toward higher specificity. Similarly, the range of CRISPR-targetable sequences could be expanded through the use of homologs with different PAM requirements (*9*), or directed evolution. Finally, inactivating one of the Cas9 nuclease domains increases the ratio of HR to NHEJ and may reduce toxicity (figs. S1A and fig. S3) (*4*, *5*), while inactivating both domains may enable Cas9 to function as a retargetable DNA binding protein. As we explore these areas, we note that another parallel study (*21*) has independently confirmed the high efficiency of CRISPR-mediated gene targeting in mammalian cell lines. We expect RNA-guided genome targeting to have broad implications for synthetic biology (*22*, *23*), the direct and multiplexed perturbation of gene networks (*13*, *24*), and targeted ex vivo (*25*–*27*) and in vivo gene therapy (*28*).

References and Notes

1. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331 (2012). doi:10.1038/nature10886 Medline
2. D. Bhaya, M. Davison, R. Barrangou, CRISPR-Cas systems in bacteria and archaea: Versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45**, 273 (2011). doi:10.1146/annurev-genet-110410-132430 Medline
3. M. P. Terns, R. M. Terns, CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**, 321 (2011). doi:10.1016/j.mib.2011.03.005 Medline
4. M. Jinek *et al*., A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816 (2012). doi:10.1126/science.1225829 Medline
5. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579 (2012). doi:10.1073/pnas.1208507109 Medline
6. R. Sapranauskas *et al*., The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res.* **39**, 9275 (2011). doi:10.1093/nar/gkr606 Medline
7. T. R. Brummelkamp, R. Bernards, R. Agami, A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550 (2002). doi:10.1126/science.1068999 Medline
8. M. Miyagishi, K. Taira, U6 promoter-driven siRNAs with four uridine 3′ overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat. Biotechnol.* **20**, 497 (2002). doi:10.1038/nbt0502-497 Medline
9. E. Deltcheva *et al*., CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602 (2011). doi:10.1038/nature09886 Medline
10. J. Zou, P. Mali, X. Huang, S. N. Dowey, L. Cheng, Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* **118**, 4599 (2011). doi:10.1182/blood-2011-02-335554 Medline
11. N. E. Sanjana *et al*., A transcription activator-like effector toolbox for genome engineering. *Nat. Protoc.* **7**, 171 (2012). doi:10.1038/nprot.2011.431 Medline
12. J. H. Lee *et al*., A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. *PLoS Genet.* **5**, e1000718 (2009). doi:10.1371/journal.pgen.1000718 Medline
13. D. Hockemeyer *et al*., Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851 (2009). doi:10.1038/nbt.1562 Medline
14. S. Kosuri *et al*., Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nat. Biotechnol.* **28**, 1295 (2010). doi:10.1038/nbt.1716 Medline
15. V. Pattanayak, C. L. Ramirez, J. K. Joung, D. R. Liu, Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* **8**, 765 (2011). doi:10.1038/nmeth.1670 Medline
16. N. M. King, O. Cohen-Haguenauer, En route to ethical recommendations for gene transfer clinical trials. *Mol. Ther.* **16**, 432 (2008). doi:10.1038/mt.2008.13 Medline
17. Y. G. Kim, J. Cha, S. Chandrasegaran, Hybrid restriction enzymes: Zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 1156 (1996). doi:10.1073/pnas.93.3.1156 Medline
18. E. J. Rebar, C. O. Pabo, Zinc finger phage: Affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671 (1994). doi:10.1126/science.8303274 Medline
19. J. Boch *et al*., Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509 (2009). doi:10.1126/science.1178811 Medline
20. M. J. Moscou, A. J. Bogdanove, A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009). doi:10.1126/science.1178817 Medline
21. L. Cong *et al*., Multiplex genome engineering using CRISPR/Cas systems. *Science* 10.1126/science.1231143 (2013).
22. A. S. Khalil, J. J. Collins, Synthetic biology: Applications come of age. *Nat. Rev. Genet.* **11**, 367 (2010). doi:10.1038/nrg2775 Medline
23. P. E. Purnick, R. Weiss, The second wave of synthetic biology: From modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410 (2009). doi:10.1038/nrm2698 Medline
24. J. Zou *et al*., Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell* **5**, 97 (2009). doi:10.1016/j.stem.2009.05.023 Medline
25. N. Holt *et al*., Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.* **28**, 839 (2010). doi:10.1038/nbt.1663 Medline
26. F. D. Urnov *et al*., Highly efficient endogenous human gene correction using

designed zinc-finger nucleases. *Nature* **435**, 646 (2005). doi:10.1038/nature03556 Medline
27. A. Lombardo *et al*., Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.* **25**, 1298 (2007). doi:10.1038/nbt1353 Medline
28. H. Li *et al*., In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217 (2011). doi:10.1038/nature10177 Medline
29. K. S. Makarova *et al*., Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467 (2011). doi:10.1038/nrmicro2577 Medline
30. P. Horvath, R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167 (2010). doi:10.1126/science.1179555 Medline
31. H. Deveau *et al*., Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390 (2008). doi:10.1128/JB.01412-07 Medline
32. J. R. van der Ploeg, Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* **155**, 1966 (2009). doi:10.1099/mic.0.027508-0 Medline
33. M. Rho, Y. W. Wu, H. Tang, T. G. Doak, Y. Ye, Diverse CRISPRs evolving in human microbiomes. *PLoS Genet.* **8**, e1002441 (2012). doi:10.1371/journal.pgen.1002441 Medline
34. D. T. Pride *et al*., Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome Res.* **21**, 126 (2011). doi:10.1101/gr.111732.110 Medline
35. J. E. Garneau *et al*., The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67 (2010). doi:10.1038/nature09523 Medline
36. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499 (2011). doi:10.1038/nature09929 Medline
37. R. Barrangou, P. Horvath, CRISPR: New horizons in phage resistance and strain identification. *Annu. Rev. Food Sci. Technol.* **3**, 143 (2012). doi:10.1146/annurev-food-022811-101134 Medline
38. W. J. Kent *et al*., The human genome browser at UCSC. *Genome Res.* **12**, 996 (2002). Medline
39. T. R. Dreszer *et al*., The UCSC Genome Browser database: Extensions and updates 2011. *Nucleic Acids Res.* **40**, (Database issue), D918 (2012). doi:10.1093/nar/gkr1055 Medline
40. D. Karolchik *et al*., The UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* **32**, (Database issue), D493 (2004). doi:10.1093/nar/gkh103 Medline
41. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841 (2010). doi:10.1093/bioinformatics/btq033 Medline
42. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi:10.1186/gb-2009-10-3-r25 Medline
43. R. Lorenz *et al*., ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011). doi:10.1186/1748-7188-6-26 Medline
44. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911 (1999). doi:10.1006/jmbi.1999.2700 Medline
45. R. E. Thurman *et al*., The accessible chromatin landscape of the human genome. *Nature* **489**, 75 (2012). doi:10.1038/nature11232 Medline
46. Q. Xu, M. R. Schlabach, G. J. Hannon, S. J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 2289 (2009). doi:10.1073/pnas.0812506106 Medline

Acknowledgments: This work was supported by NIH grant P50 HG005550. We thank Sriram Kosuri for advice on the oligonucleotide pool designs and synthesis. G.M.C. and P.M. have filed a patent based on the findings of this study.

Supplementary Materials
www.sciencemag.org/cgi/content/full/science.1232033/DC1
Materials and Methods
Figs. S1 to S11
Tables S1 to S3

References (*29*–*46*)

26 October 2012; accepted 12 December 2012
Published online 03 January 2013; 10.1126/science.1232033

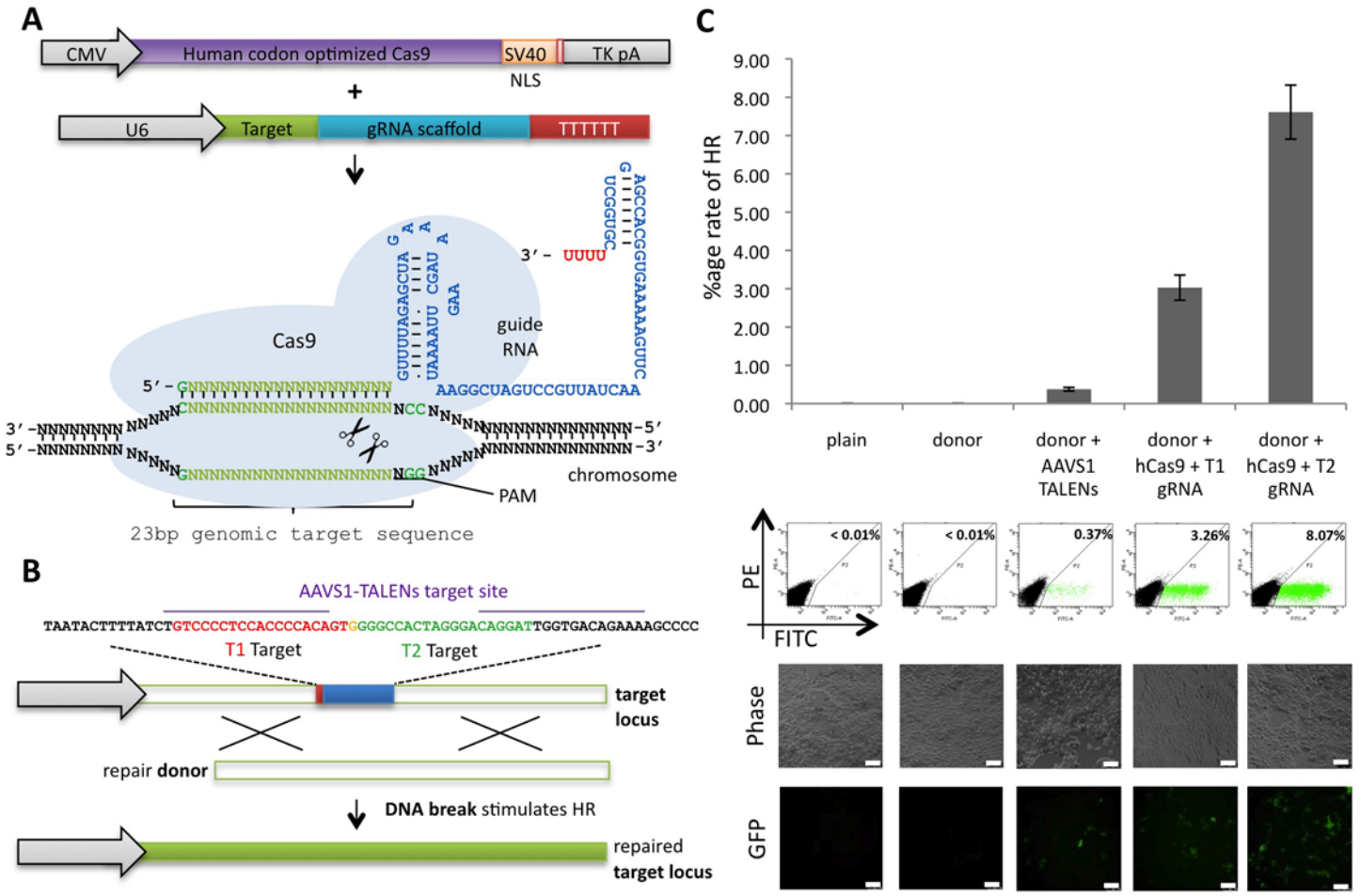

Fig. 1. Genome editing in human cells using an engineered type II CRISPR system. (**A**) RNA-guided gene targeting in human cells involves co-expression of the Cas9 protein bearing a C terminus SV40 nuclear localization signal with one or more guide RNAs (gRNAs) expressed from the human U6 polymerase III promoter. Cas9 unwinds the DNA duplex and cleaves both strands upon recognition of a target sequence by the gRNA, but only if the correct protospacer-adjacent motif (PAM) is present at the 3′ end. Any genomic sequence of the form $GN_{20}GG$ can in principle be targeted. (**B**) A genomically integrated GFP coding sequence is disrupted by the insertion of a stop codon and a 68bp genomic fragment from the AAVS1 locus. Restoration of the GFP sequence by homologous recombination (HR) with an appropriate donor sequence results in GFP^{+} cells that can be quantitated by FACS. T1 and T2 gRNAs target sequences within the AAVS1 fragment. Binding sites for the two halves of the TAL effector nuclease heterodimer (TALEN) are underlined. (**C**) Bar graph depicting HR efficiencies induced by T1, T2, and TALEN-mediated nuclease activity at the target locus, as measured by FACS. Representative FACS plots and microscopy images of the targeted cells are depicted below (scale bar is 100 microns). Data are means +/− SEM (N=3).

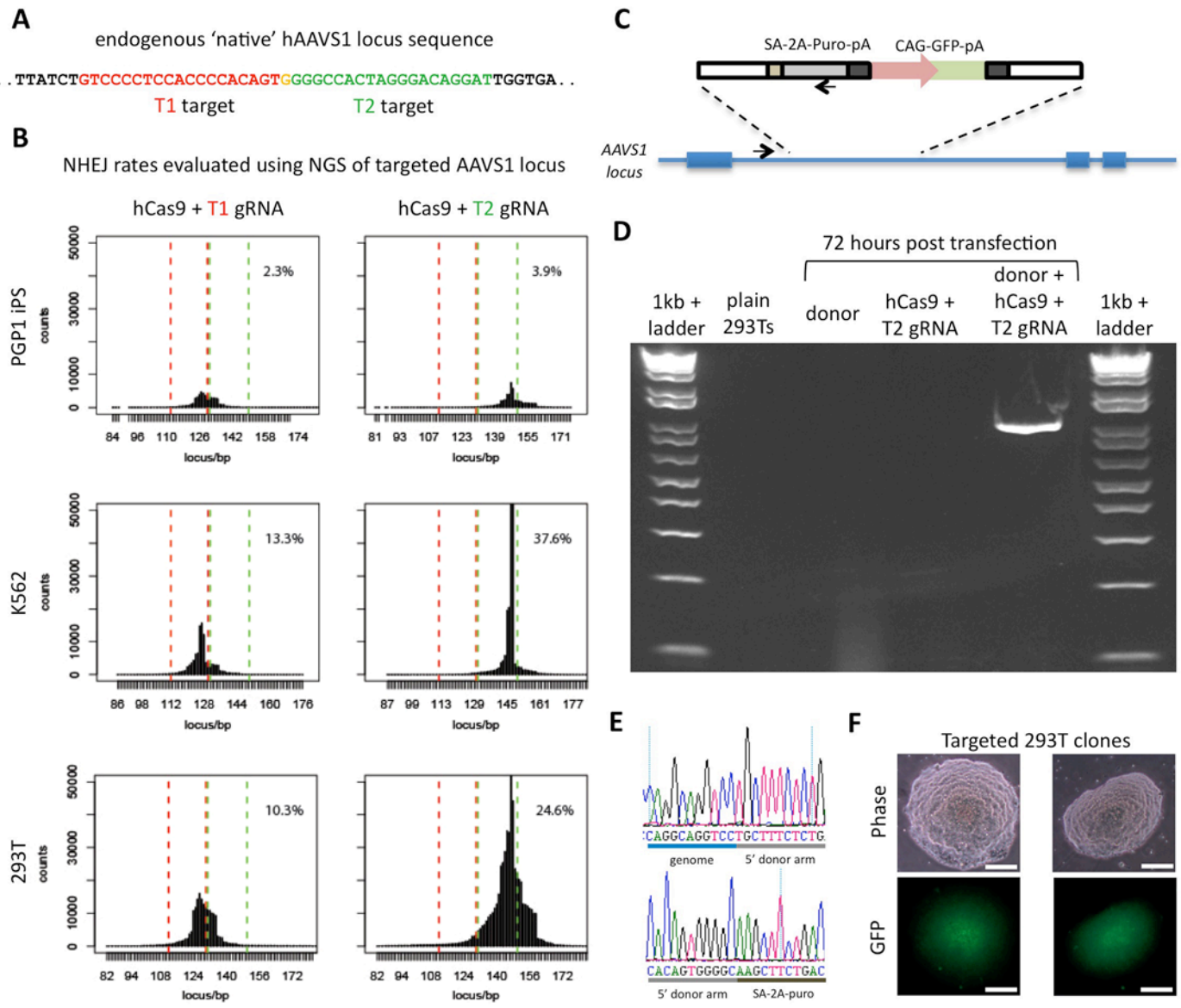

Fig. 2. RNA-guided genome editing of the native AAVS1 locus in multiple cell types. (**A**) T1 (red) and T2 (green) gRNAs target sequences in an intron of the PPP1R12C gene within the chromosome 19 AAVS1 locus. (**B**) Total count and location of deletions caused by NHEJ in 293Ts, K562s, and PGP1 iPS cells following expression of Cas9 and either T1 or T2 gRNAs as quantified by next-generation sequencing. Red and green dash lines demarcate the boundaries of the T1 and T2 gRNA targeting sites. NHEJ frequencies for T1 and T2 gRNAs were 10% and 25% in 293T, 13% and 38% in K562, and 2% and 4% in PGP1 iPS cells, respectively. (**C**) DNA donor architecture for HR at the AAVS1 locus, and the locations of sequencing primers (arrows) for detecting successful targeted events, are depicted. (**D**) PCR assay three days post transfection demonstrates that only cells expressing the donor, Cas9 and T2 gRNA exhibit successful HR events. (**E**) Successful HR was confirmed by Sanger sequencing of the PCR amplicon showing that the expected DNA bases at both the genome-donor and donor-insert boundaries are present. (**F**) Successfully targeted clones of 293T cells were selected with puromycin for 2 weeks. Microscope images of two representative GFP+ clones is shown (scale bar is 100 microns).