Precise Cas9 targeting enables genomic mutation prevention

Alejandro Chavez\textsuperscript{a,b,c,1,2}, Benjamin W. Pruitt\textsuperscript{b,1}, Marcelle Tuttle\textsuperscript{b,1}, Rebecca S. Shapiro\textsuperscript{b,d,e,f,g,3}, Ryan J. Cecchi\textsuperscript{b,1}, Jordan Winston\textsuperscript{b,1}, Brian M. Turczyk\textsuperscript{b,6}, Michael Tung\textsuperscript{b,9}, James J. Collins\textsuperscript{b,d,e,f,g,3}, and George M. Church\textsuperscript{b,c,2}

\textsuperscript{a}Department of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, NY 10032; \textsuperscript{b}Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA 02115; \textsuperscript{c}Department of Genetics, Harvard Medical School, Boston, MA 02115; \textsuperscript{d}Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA 02139; \textsuperscript{e}Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA 02139; \textsuperscript{f}Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA; and \textsuperscript{g}Broad Institute of MIT and Harvard, Cambridge, MA 02142

Contributed by George M. Church, January 30, 2018 (sent for review October 27, 2017; reviewed by Yannick Doyon and Junwei Shi)

Here, we present a generalized method of guide RNA “tuning” that enables Cas9 to discriminate between two target sites that differ by a single-nucleotide polymorphism. We employ our methodology to generate an in vivo mutation prevention system in which Cas9 actively restricts the occurrence of undesired gain-of-function mutations within a population of engineered organisms. We further demonstrate that the system is scalable to a multitude of targets and that the general tuning and prevention concepts are portable across engineered Cas9 variants and Cas9 orthologs. Finally, we show that the mutation prevention system maintains robust activity even when placed within the complex environment of the mouse gastrointestinal tract.

Cas9 | tuned gRNA | tgRNA | mutation prevention | synthetic biology

T

here are few biological perturbations that rival point mutations when it comes to the power to affect phenotypic change. Single-base substitutions endow pathogens with resistance to antibiotics and rogue cells with oncogenic potential (1, 2). Despite decades of research into the causes and effects of point mutations, no tool exists to directly prevent their occurrence. At best, we can track when mutations occur and utilize them as prognostic factors to predict emergent properties or chemotherapeutic outcome (3, 4). Here we describe an in vivo “mutation prevention” system that can prevent the occurrence of targeted point mutations with little to no latent toxicity to the host organism.

The system is predicated on Streptococcus pyogenes Cas9 (hereon referred to as Cas9) and its orthologs: endonucleases directed to a target locus via hybridization between an associated guide RNA (gRNA) and a target site near a required protospacer adjacent motif (PAM) (5–8). While incredibly plastic, Cas9 suffers from difficult-to-predict nonspecific activity, which can be tolerant to multiple mismatches between the gRNA and the inappropriately bound locus (9–20). Several groups have presented approaches that confer greater specificity to Cas9 (in some cases demonstrating single-nucleotide specificity) (13, 21–29). These techniques, including those involving engineered Cas9 proteins, can still fall short in applications where near-absolute single-nucleotide discrimination is required, necessitating the introduction of alternative approaches. To endow Cas9 with the ability to discriminate between single-nucleotide polymorphisms (SNPs), we developed a screening methodology that confers single-nucleotide specificity through the selection of a tuned guide RNA (tgRNA).

Results

To demonstrate the feasibility of preventing the emergence of point mutations with a Cas9/tgRNA system, we generated a strain of Escherichia coli deficient in mismatch repair (MG1655-mutS::kan), to increase mutation rates, and harboring a plasmid encoding a catalytically inactive version of the TEM-1 β-lactamase (TEM-1-S68N) (30, 31). The active version of TEM-1 confers resistance to β-lactam antibiotics, such as ampicillin (32). Under typical growth conditions, errors in DNA replication and repair lead to the accumulation of mutations over time. Some of these mutations result in the reversion of the inactive N68 allele to its catalytically active form, S68. The number of revertants within the population is quantified by plating overnight cultures to solid media containing ampicillin (Fig. 1A).

We postulated that we could prevent these reversions by introducing a Cas9 system tuned to cut the active TEM-1 variant while exhibiting undetectable activity against the inactive allele. In practice, such a system would benignly persist within cells until

Significance

Single-base substitutions are capable of producing transformative phenotypic changes. While methods to classify such mutations are well established, it is difficult to modulate or preclude their occurrence in a direct and efficacious manner. In this study, we refine the specificity of the CRISPR-Cas9 system and present a general framework for proactively preventing the occurrence of point mutations. This “mutation prevention system” is a broadly useful tool for the study and control of DNA substitutions, particularly in contexts where an associated phenotype or evolutionary pathway is undesirable.


Reviewers: Y.D., Université Laval; and J.S., University of Pennsylvania.

Conflict of interest statement: G.M.C. is a founder and advisor for Editas Medicine. G.M.C. has equity in Editas, Enevol, and Caribou Biosciences. A patent has been filed by Harvard University related to the work within this article.

Published under the PNAS license.

Data deposition: The sequences reported in this paper have been deposited with the National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/clinvar/PRJNA329016 (accession nos. SRR3909289, SRR3909312–SRR3909322, SRR3909346, SRR3909347, SRR3909359, SRR3909371–SRR3909373, SRR3909397–SRR3909399, SRR3909417, SRR3909432, SRR3909435, SRR3909456, SRR3909457, SRR3909458, SRR3909460, SRR3909483, SRR3909484, SRR3909507, SRR3909508, SRR3909531, SRR3909532, SRR3909553, SRR3909554, SRR3909578, SRR3909579, SRR3909580, SRR3909589, SRR3909603, SRR3909621, SRR3909634, SRR3909654, SRR3909655, SRR3909694, SRR3909695, SRR3909711, SRR3909719, SRR3909731, SRR3909747, SRR3909770, SRR3909786, SRR3909795, SRR3909813, SRR3909820, SRR3909821, SRR3909844, SRR3909848, SRR3909866, SRR3909887, SRR3909892, SRR3909907, SRR3909923, SRR3909924, SRR3909949, SRR3909950, SRR3909953, SRR3909974, SRR3909997, SRR3909998, SRR3910021, SRR3910022, SRR3910045, SRR3910062, SRR3910072, SRR3910084, SRR3910085, SRR3910086, SRR3910087, SRR3910088, SRR3910090, SRR3910091, SRR3910092, SRR3910093, SRR3910094, SRR3910095, SRR3910096, SRR3910117, SRR3910118, SRR3910129, and SRR3910139.

\textsuperscript{1}A.C. and B.W.P. contributed equally to this work.

\textsuperscript{2}To whom correspondence may be addressed. Email: ac4304@cumc.columbia.edu or gchurch@genetics.med.harvard.edu.

\textsuperscript{3}Present address: Department of Molecular and Cellular Biology, University of Guelph, Guelph, ON N1G2W1, Canada.

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1718148115/-/DCSupplemental.

Published online March 19, 2018.

www.pnas.org/cgi/doi/10.1073/pnas.1718148115

PNAS | April 3, 2018 | vol. 115 | no. 14 | 3669–3673
the occurrence of the reversion. Cas9 would then cut at the active allele, causing irreparable genetic damage and a decrease in the overall number of ampicillin-resistant cells (Fig. 1B). By design, the inactive form of TEM-1 differs from the active form by a single-nucleotide substitution (203G > A). Our initial, naïve system included a gRNA with full complementarity to the active S68 allele (203G) and thus only differed from the inactive N68 allele by a single nucleotide. As expected, this approach performed poorly, as the Cas9/gRNA system cut both the active and inactive forms efficiently. Consequently, the system exhibited high toxicity even in the absence of the reversion (data not shown). Furthermore, we found that even the high-fidelity eCas9 and Cas9-HF1 variants failed to confer single-nucleotide discrimination with the naïve one-off gRNA, as indicated by a sharp decrease in viability for cells carrying the system (25, 26) (SI Appendix, Table S1).

Previous work examining the factors that influence Cas9 behavior has shown that its activity is easily modulated by the introduction of mismatches within the gRNA that prevent full complementarity between it and the target site (5, 10, 11). We posited that, for a given pair of targets that differ by a single base, there may exist a set of mismatches within the gRNA that would eliminate activity on one of the variants while maintaining robust activity on the other. To test our hypothesis, we screened a library of gRNAs that each differed from the active TEM-1 allele (S68) by a single mismatch and from the inactive TEM-1 allele (N68) by two mismatches (SI Appendix, Fig. S1). We focused our mutational analysis on the region within the gRNA that interacts with the bases proximal to the PAM at the target locus, as these residues have been shown to be most critical for Cas9 activity (5, 9–11).

Our screen was designed such that an ideal tgRNA candidate was expected to be markedly depleted when in the presence of the functional S68 allele (signifying that it cut the undesired TEM-1 variant that we wish to prevent) and strongly enriched when tested against the nonfunctional N68 allele (suggesting that it did not cut the desired TEM-1 variant we wish to maintain). Upon performing the screen, we identified several promising tgrNA candidates that were subjected to further characterization (Fig. 1C).

Individual testing of these select library members revealed that the fold depletion we observe within the screen strongly correlates with their respective activities in isolation (SI Appendix, Fig. S2). This validates the capability of our library approach to accurately identify the most discriminatory gRNAs. Based on the results of our screen and subsequent validation, we selected the three candidate tgrNA with the greatest discriminatory power, 6A, 6G, and 6T, for additional analysis. We tested their ability to prevent emergence of the active TEM-1 allele from within a population of cells containing only the inactive TEM-1 variant. As predicted, all three tgrNAs were able to prevent reversion to the active TEM-1 allele by several orders of magnitude over cells with a nonfunctional control gRNA, while also exhibiting minimal levels of baseline toxicity (Fig. 1 D and E).

Because SP-Cas9 represents one of thousands of known Cas9 proteins, each with differences in efficiency, specificity, and PAM requirements, we sought to verify that our approach was generalizable to two commonly used Cas9 orthologs: NM-Cas9 and ST1-Cas9.
We designed mutation prevention systems with gRNAs targeting rpsL and rpoB genes. A gRNA designed to prevent the rpsL.1534T>C mutation, we screened a series of gRNAs and identified a canonical gRNA, bearing additional mutations, that was highly efficient (Fig. 2B). Targeting the rpoB.1547A>G mutation was simplified by the fact that the mutation generates a PAM that is not present within the wild-type rpoB gene. A gRNA designed to bind in the presence of the generated PAM conferred marked mutation prevention (Fig. 2B). Finally, to prevent the rpoB.1546G>T substitution, we expressed a pair of gRNAs (denoted as dual gRNA) along with an appropriate gRNA to prevent the mutation. With rpoB.1546G>T, the mutation generates a PAM that induces toxicity to nonmutant cells (Fig. 2B).

Rifampicin is a widely used antibiotic that inhibits RNA polymerase function, with resistance arising from several well-documented mutations within the rpoB gene (38, 39). To further demonstrate the plasticity of our approach, we targeted a series of the prominent rifampicin-resistance mutations and generated unique gRNAs against each using three disparate design principles. To prevent the rpoB.1534T>C mutation, we screened a series of gRNAs and identified a canonical gRNA, bearing additional mutations, that was highly efficient (Fig. 2B). Targeting the rpoB.1547A>G mutation was simplified by the fact that the mutation generates a PAM that is not present within the wild-type rpoB gene. A gRNA designed to bind in the presence of the generated PAM conferred marked mutation prevention (Fig. 2B). Finally, to prevent the rpoB.1546G>T substitution, we expressed a pair of gRNAs (denoted as dual gRNA) along with an appropriate gRNA to prevent the mutation. With rpoB.1546G>T, the mutation generates a PAM that induces toxicity to nonmutant cells (Fig. 2B).
mutation, we employed the Cas9-VQR variant with altered PAM specificity (40). This was necessary given the lack of a canonical PAM sequence near the mutation site. Cas9-VQR, in the presence of an appropriate gRNA, led to efficient mutation prevention (Fig. 2B).

Having demonstrated both the portability and flexibility of our mutation prevention concept, we next assessed its scalability by targeting multiple mutations within *rpoB* with a single system. There was no obvious decrease in efficiency upon multiplexing, with both *rpoB* 1534T > C and 1547A > G mutations being simultaneously prevented within the engineered strain (Fig. 2C).

To further probe the limits of our multimutation prevention system, we sought to simultaneously inhibit all of the previously targeted endogenous mutations (except *rpoB* 1546G > T, which requires the Cas9-VQR variant) within a single strain. Cells carrying a full complement of four tgRNAs exhibited minimal toxicity compared with cells carrying nontargeting control gRNAs and retained robust mutation prevention against all targeted loci (Fig. 2D and SI Appendix, Table S2).

We next wanted to assess the robustness of our system by preventing point mutations from emerging within a complex in vivo environment. Toward this goal we inoculated gnotobiotic mice with one of two strains containing either a pair of nontargeted gRNAs or a pair of tgRNAs directed to prevent mutations that confer rifampicin resistance within *E. coli*. Two days after colonization, mice were provided rifampicin within their drinking water and the spectrum of rifampicin resistance mutations was analyzed in bacterial cells recovered from fecal samples collected on days 4–7, postcolonization (Fig. 3A). We observed minimal frequencies of the targeted point mutations within the engineered strain, relative to the much higher frequencies within the control, across all days (Fig. 3B).

**Discussion**

It is expected that the mutator strain of *E. coli* utilized within all of our experiments would commensurately increase the frequency of mutations that disable the mutation prevention system itself (3, 30). Given that the system is likely breaking at an appreciable rate, the fact that we do not see obvious degradation in its performance over time likely indicates that the system places a negligible fitness burden on the host organism. If the system were to place a significant fitness burden on its host, one would expect individuals with attenuated or broken systems to rapidly overtake the population, thus facilitating escape from its selection. It is thanks to this minimal fitness burden that escape is an extremely rare event. This is due to the fact that during any given point in time there is only a small population of cells within the total population that have inactivated the system and it is within this rare population that a second targeted gain-of-function mutation must occur. This “two-hit” requirement effectively makes escape from our mutation prevention strategy an extremely rare event and explains our efficacy at preventing undesired mutations.

The robust performance of the mutation prevention system within the mouse gastrointestinal tract is particularly notable given its tolerance of multiday rifampicin selection, the absence of the active selection for the episomal plasmid encoding our system, and the fact that these experiments were performed within a highly mutagenic mismatch repair deficient background of *E. coli*.

Although our mutation prevention strategy requires antecedent knowledge of the target mutation(s), we note that there is already a plethora of known gain-of-function mutations that endow living cells with undesired phenotypes. In these cases, users may want to prevent a subset of these mutations to gain deeper insight into the existence and effects of other rarer alleles. In addition, rapid increases in both the affordability and adoption rate of next-generation sequencing will catalyze the eventual systematic characterization of genomic mutations that enable cells to evade the selective pressures imposed upon them by both natural and artificial systems. Our method will enable experimentalists to capitalize upon this expansive characterization, and in turn, manipulate it to prevent undesired outcomes.

This study focused on preventing gain-of-function mutations within microbial systems, in which cutting leads to loss of episomal elements or lethality when the target is present within the genome (41–44). When applying similar strategies to higher eukaryotes the effect of nonhomologous end joining-based mutagenesis (NHEJ) must be considered, as nonactivating mutations have been shown to occur that can themselves lead to gain-of-function phenotypes (45, 46). As the outcome of NHEJ-based repair has been shown to be nonrandom, a potential strategy to mitigate Cas9 derived gain-of-function alleles would be to prescreen the selected target locus and tgRNA to determine the spectrum of observed repair profiles, post Cas9-mediated cleavage.

The design of sufficiently discriminatory tgRNAs currently necessitates library screens in the spirit of those described within this text. That is, it is necessary to empirically test a small library of candidate tgRNAs to find the guide sequence that is both effective at cutting the target allele while exhibiting minimal cutting against the desired allele. Fortunately, these screens are easy to perform with standard laboratory equipment and do not require costly high-throughput oligo synthesis. In some cases, more extensive screens (or further system optimization) may be required due to lingering levels of system toxicity, as is evidenced by the mild decreases in cfu/mL that we saw with the NM and ST1 tgRNA libraries (SI Appendix, Fig. S3). We do not, at this time, have sufficient screening data to propose an in silico model for forward design of tgRNAs; however, we do anticipate that, given sufficient adoption of this technique, such an approach may be possible in the future.

Even in its nascent form, the described mutation prevention system provides the scientific community with a powerful means to propose an in silico model for forward design of tgRNAs; however, we do anticipate that, given sufficient adoption of this technique, such an approach may be possible in the future.
to directly manipulate evolutionary outcomes and in doing so gain fundamental insights into evolutionary plasticity and adaptive mechanisms. Furthermore, as our ability to engineer targeted nucleases such as Cas9 continues to rapidly improve, one can imagine the eventual generation of increasingly complex mutation prevention strategies. We further expect that the generalizable principle of guide RNA tuning will exhibit synergistic performance improvements with newly engineered, high-sensitivity variants of Cas9 (26, 27, 29), as well as orthogonal DNA binding proteins like Cpf1 (47, 48). As a result, applications requiring true single-nucleotide specificity, such as the disabling of diseased alleles delineated by SNPs, will finally be realized.

Materials and Methods

Experiments were performed in an E. coli MG1655-mutS::kan background (received as a gift from Dr. Harris Wang, Columbia University, New York); all gRNA/tgRNA screening experiments were performed in a wild-type E. coli MG1655 background. Overnight liquid cultures were grown in lysogeny broth (LB) supplemented with the appropriate antibiotic(s) and incubated at 37 °C for 12 h, with shaking. Overnight solid cultures (including spot assays and mutation assays) were grown on LB agar plates supplemented with the appropriate antibiotic(s) and incubated at 37 °C for 12–16 h. The mutation rates for experiments were determined using either resistant colony-forming units (TEM-1 based assay) or through next generation sequencing (endogenous mutation prevention experiments). All mouse experiments were performed in accordance with and as approved by the Harvard Medical School Institutional Animal Care and Use Committee (protocol no. 04957) and Harvard Medical School Committee on Microbiological Safety (approval no. 12-085). Detailed protocols are contained within SI Appendix.

ACKNOWLEDGMENTS. We thank L. Certain, X. Guo, A. Tsou, A. Tung, S. Vora, N. C. Yeo, and J. Zhu for helpful discussions and technical assistance. We acknowledge research support from the US National Institutes of Health National Human Genome Research Institute Grant PS0 HG005550, and the Wyss Institute for Biologically Inspired Engineering. A.C. was funded by the National Cancer Institute Grant ST32CA092164-34 and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. J.J.C. was supported by the Defense Threat Reduction Agency Grant HDTRA1-15-1-0051 and the Paul G. Allen Frontiers Groups.