Orthogonal Cas9 proteins for RNA-guided gene regulation and editing

Kevin M Esvelt[1,4], Prashant Mali[2,4], Jonathan L Braff[1], Mark Moosburner[2], Stephanie J Yaung[1–3] & George M Church[1,2]

The Cas9 protein from the *Streptococcus pyogenes* CRISPR-Cas acquired immune system has been adapted for both RNA-guided genome editing and gene regulation in a variety of organisms, but it can mediate only a single activity at a time within any given cell. Here we characterize a set of fully orthogonal Cas9 proteins and demonstrate their ability to mediate simultaneous and independently targeted gene regulation and editing in bacteria and in human cells. We find that Cas9 orthologs display consistent patterns in their recognition of target sequences, and we identify an unexpectedly versatile Cas9 protein from *Neisseria meningitidis*. We provide a basal set of orthogonal RNA-guided proteins for controlling biological systems and establish a general methodology for characterizing additional proteins.

Clustered, regularly interspaced, short palindromic repeats (CRISPR)–CRISPR-associated (Cas) systems provide bacteria and archaea with acquired immunity by incorporating fragments of viral or plasmid DNA into CRISPR loci and using the transcribed crRNAs to guide the degradation of homologous sequences[1,2]. In type II CRISPR systems, a ternary complex of a Cas9 nuclease with CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA) binds to and cleaves dsDNA protospacer sequences that match the crRNA spacer and abut a short protospacer-adjacent motif (PAM) recognized by Cas9 (refs. 3,4). Fusing the crRNA and tracrRNA produces a single guide RNA (sgRNA) that is sufficient to target Cas9 (ref. 4).

As an RNA-guided nuclease and nickase, Cas9 has been adapted for targeted gene editing[5–9] and selection[10] in a variety of organisms. Although these successes are arguably transformative, nuclease-null Cas9 variants may prove to be at least as useful for regulatory purposes, as the ability to localize proteins and RNA to nearly any set of dsDNA sequences affords tremendous versatility for controlling biological systems[11–17]. Beginning with targeted gene repression through promoter and 5′ UTR obstruction in bacteria[18], Cas9-mediated regulation was recently extended to transcriptional activation by means of VP64 recruitment in human cells[19–22]. Looking forward, we anticipate a cornucopia of Cas9-mediated transcriptional activators, repressors, fluorescent protein labels, chromosome tethers and numerous other tools.

Although the Cas9 protein from *S. pyogenes* can mediate one activity at many different target sites, it cannot concurrently mediate a different activity at other targets. For example, a cell engineered with a Cas9 activator cannot undergo genome editing using a Cas9 nuclease without it also cutting the sites being targeted by the activator. Simultaneously employing multiple RNA-guided activities within a single cell will require methods of independently targeting each activity to its own set of target sites. To establish this level of control[23,24], we developed methods enabling the characterization of orthogonal Cas9 proteins for multiplexed RNA-guided transcriptional activation, repression and gene editing.

RESULTS

Selecting putatively orthogonal Cas9 proteins

Cas9 RNA binding and sgRNA specificity is primarily determined by the ~36-bp repeat sequence in pre-crRNA. We began by examining known Cas9 genes for highly divergent repeats in their adjacent CRISPR loci. We chose the genes encoding the well-studied Cas9 protein from *S. pyogenes* (SP), the smaller Cas9 proteins from *Streptococcus thermophilus* CRISPR1 and *N. meningitidis* (ST1 and NM, respectively) and the large Cas9 protein from *Treponema denticola* (TD). The CRISPR loci associated with these genes harbor repeats that differ by at least 13 nucleotides from one another (**Fig. 1a**).

Protospacer-adjacent–motif characterization

Known Cas9 proteins will target only dsDNA sequences flanked by a 3′ PAM sequence specific to the Cas9 of interest. Of the four Cas9 variants, only SP has an experimentally characterized PAM, whereas the ST1 PAM and, very recently, the NM PAM were deduced bioinformatically. SP is thought to be the most readily targetable of these owing to its short PAM of NGG[10]; ST1 and NM targeting are constrained by PAMs of NNAGAAW and NNNNGATT, respectively[25,26]. We hypothesized that, because

[1]Wyss Institute for Biologically Inspired Engineering, Harvard Medical School, Boston, Massachusetts, USA. [2]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [3]Program in Medical Engineering & Medical Physics, Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to G.M.C. (gchurch@genetics.med.harvard.edu).

RECEIVED 20 JUNE; ACCEPTED 27 AUGUST; PUBLISHED ONLINE 29 SEPTEMBER 2013; DOI:10.1038/NMETH.2681

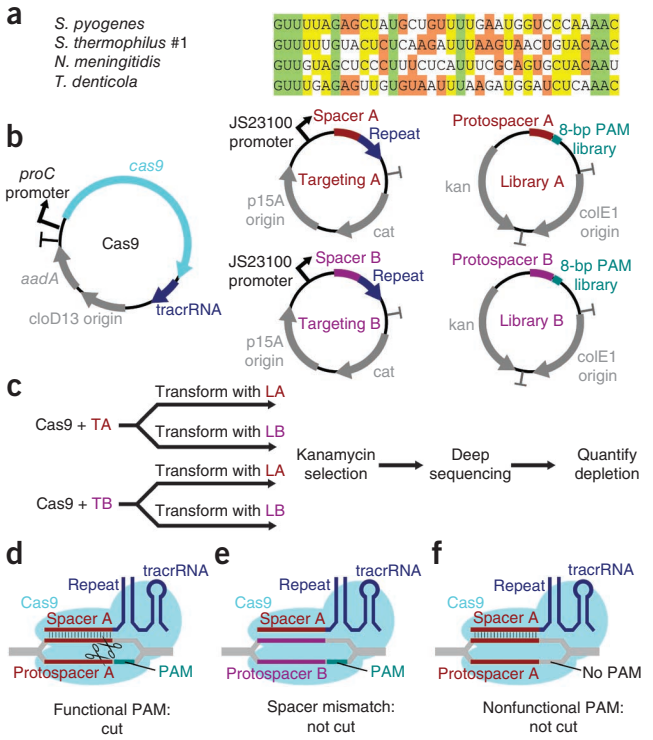

Figure 1 | Comparison and characterization of putatively orthogonal Cas9 proteins. (**a**) Repeat sequences of SP, ST1, NM and TD. Bases are colored to indicate the degree of conservation. (**b**) Plasmids used for characterization of Cas9 proteins in *E. coli*. All carry compatible replication origins and antibiotic-resistance genes. (**c**) Selection scheme to identify PAMs. Cells expressing a Cas9 protein and one of two spacer-containing targeting plasmids (TA or TB) were transformed with one of two PAM libraries (LA or LB) with corresponding protospacers and subjected to antibiotic selection. Surviving uncleaved plasmids were subjected to deep sequencing, and Cas9-mediated PAM depletion was quantified. (**d**–**f**) Functional PAMs are depleted from the library by Cas9 when the targeting plasmid spacer matches the library plasmid protospacer (**d**); Cas9 does not cut when the spacer and protospacer do not match (**e**); and nonfunctional PAMs are never cut or depleted (**f**).

of the additional requirement for spacer acquisition in natural systems, bioinformatic approaches might infer more stringent PAM requirements for Cas9 activity than are empirically necessary for effector cleavage. Because the PAM sequence is the most frequent target of mutation in phages that have escaped immunity, greater stringency during spacer acquisition might provide redundancy and sometimes preclude evolutionary escape. We therefore adopted a library-based approach to comprehensively characterize these sequences in bacteria via high-throughput sequencing.

Genes encoding ST1, NM and TD were assembled from synthetic fragments and cloned into bacterial expression plasmids with their associated tracrRNAs (**Fig. 1b** and **Supplementary Fig. 1**). Prior experience with variably effective spacer sequences using SP (data not shown) led us to select two SP functional spacers for incorporation into the six targeting plasmids. Each targeting plasmid encodes a constitutively expressed crRNA in which one of the two spacers is followed by the 36-bp repeat sequence specific to a Cas9 protein (**Fig. 1b**). Plasmid libraries containing one of the two protospacers followed by all possible 8-bp PAM sequences were generated by PCR and assembly[27]. Future experiments could use sequences of >8 bp in libraries to account for even longer PAMs. Each library was electroporated into *Escherichia coli* harboring Cas9 expression and targeting plasmids, for a total of 12 combinations of Cas9 protein, spacer and protospacer (**Fig. 1c**). Surviving library plasmids were selectively amplified by barcoded PCR and sequenced by MiSeq to distinguish between functional PAM sequences, which are depleted only when the spacer and protospacer match (**Fig. 1d**,**e**), from nonfunctional PAMs, which are never depleted (**Fig. 1f**).

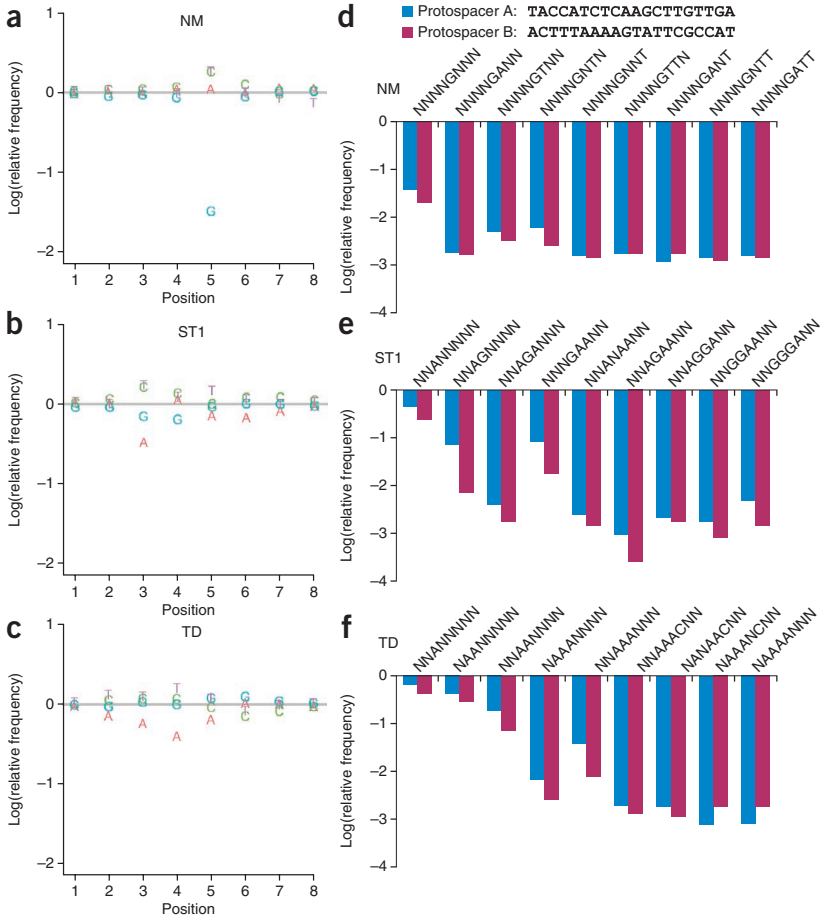

Figure 2 | Depletion of functional PAMs from libraries by Cas9 proteins. (**a**–**c**) Log frequency of each base at every position for matched spacer-protospacer pairs plotted relative to control conditions in which spacer and protospacer did not match. The results reflect the mean depletion of libraries by NM (**a**), ST1 (**b**) and TD (**c**) on the basis of two distinct protospacer sequences (A and B). (**d**–**f**) Depletion of specific sequences for each protospacer plotted separately for each Cas9 protein.

Table 1 | Protospacer-adjacent motifs for each Cas9

NM	ST1	TD
NNNN**GA**NN	NNA**GAA**	N**AAAA**N
NNNN**GTT**N	NN**AGGA**	N**AAA**N**C**
NNNN**G**NN**T**	NN**GGAA**	N**A**N**AAC**
NNNN**GT**NN	NN**A**N**AA**	NN**AAAC**
NNNN**G**N**T**N	NN**GGGA**	

Two thresholds of activity were defined for PAMs: a moderate (not underlined) and a stringent threshold (underlined).

To graphically depict the importance of each nucleotide at every position, we plotted the logarithm of the frequency of each base for matched spacer-protospacer pairs relative to the corresponding mismatched case (**Fig. 2**). As hypothesized, our results revealed that NM and ST1 recognized PAMs that are less stringent and more complex than indicated by earlier bioinformatic predictions, suggesting that requirements for spacer acquisition are indeed more stringent than those for effector cleavage. Most strikingly, NM absolutely required a single G nucleotide positioned five bases from the 3′ end of the protospacer (**Fig. 2a**), whereas ST1 and TD each required at least three specific bases (**Fig. 2b**,**c**). Sorting our results by position allowed us to quantify depletion of any PAM sequence from each protospacer library (**Fig. 2d**–**f**). All three enzymes cleaved protospacer B more effectively than protospacer A when paired with most PAMs, with ST1 exhibiting the greatest preference (**Supplementary Fig. 2a**). However, there was also considerable PAM-dependent variation in this interaction. For example, NM cleaved protospacers A and B approximately equally when they were followed by sequences matching TNNNGNNN but was tenfold more active in cleaving protospacer B for the set of sequences with PAMs matching ANNNGNNN (**Supplementary Fig. 2b**).

Our results highlight the difficulty of defining a single acceptable PAM for a given Cas9 (**Supplementary Note 1**). Not only did activity levels depend upon the sequence of the protospacer, but specific combinations of unfavorable PAM bases substantially reduced activity even when the primary base requirements were met. We initially identified PAMs as patterns that underwent >100-fold average depletion with the lower-activity protospacer A and >50-fold depletion of all derivative sequences in which one base unspecified in the parent (for example, N) was set to A, T, C or G. (**Table 1**). Although these levels are presumably sufficient to defend against targets in bacteria, we noticed that particular combinations of deleterious mutations dramatically reduced activity. For example, NM depleted sequences matching NCCAGGTN by only fourfold (PAM matches underlined; **Supplementary Fig. 2c**). We therefore defined a more stringent threshold requiring >500-fold depletion of matching sequences and >200-fold depletion of one-base derivatives for applications requiring higher affinity (**Table 1**).

Orthogonality in bacteria

We originally selected our set of Cas9 proteins for their disparate crRNA repeat sequences. To verify that they were indeed orthogonal, we cotransformed each Cas9 expression plasmid with each of the four targeting plasmids containing spacer B. These cells were challenged by transformation with substrate plasmids containing either protospacer A or protospacer B and a suitable PAM. We observed plasmid depletion exclusively when each Cas9 was paired with its own crRNA, demonstrating that the four constructs are orthogonal in bacteria (**Fig. 3**).

Transcriptional regulation in bacteria

A nuclease-null variant of SP has been demonstrated to repress targeted genes in bacteria with an efficacy dependent upon the position of the targeted protospacer[18,20,28]. We asked whether nuclease-null variants of the new proteins might be similarly capable of targeted repression. We identified the catalytic residues of the RuvC and HNH nuclease domains of each ortholog by sequence homology and inactivated them to generate nuclease-null NM, ST1 and TD (Online Methods). To create suitable reporters, we inserted protospacer B with an appropriate PAM for each Cas9 into the nontemplate strand within the coding sequence of an EYFP reporter plasmid (**Fig. 4a**). We cotransformed each of these constructs into *E. coli* together with their corresponding targeting plasmids and measured the resulting fluorescence. Cells with matching spacer and protospacer exhibited much weaker fluorescence than the corresponding mismatched case for nuclease-null SP, ST1 and especially NM, but less so for TD. To determine whether this was an artifact of the low basal activity of the TD reporter, we also tested an alternative reporter design in which protospacer A was placed in the 5′ UTR, which confirmed that TD is much less effective as a repressor (**Fig. 4b**). These results indicate that not all Cas9 proteins are equally suitable for every task and suggest possible differences between larger Cas9 proteins such as TD and smaller members of the family. More practically, our results demonstrate that three of our four orthologs can function as robust RNA-guided repressors in bacteria.

Simultaneous gene regulation and nuclease activity

Having demonstrated that our orthogonal Cas9 proteins are capable of both nuclease activity and transcriptional repression, we next engineered *E. coli* to employ both activities simultaneously. We constructed a plasmid encoding SP to defend against filamentous phage infection and used our previous constructs encoding nuclease-null NM, the most readily targetable of the orthologs, to repress the EYFP reporter. As expected, the resulting cells successfully repressed EYFP transcription and cleaved incoming filamentous phage genomes at multiple locations within gene III, completely preventing plaque formation by M13mp18 (**Fig. 4c**) and precluding transformation with a plasmid containing the targeted gene (**Fig. 4d**). These results demonstrate the ability of

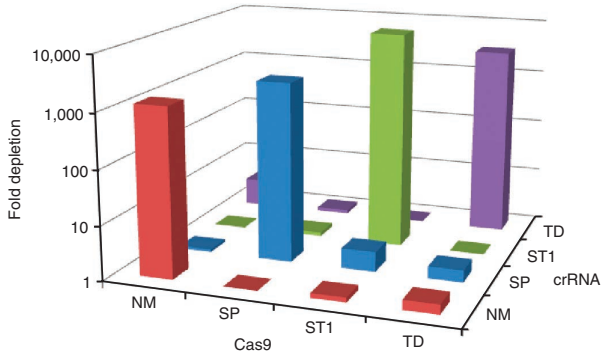

Figure 3 | Orthogonal recognition of crRNAs in *E. coli*. Cells with the indicated 16 combinations of Cas9 and crRNA were challenged with a plasmid bearing a matched or mismatched protospacer and appropriate PAM. Sufficient cells were plated to reliably obtain colonies from matching spacer and protospacer pairings. Total colony counts on the resulting 32 plates were used to calculate the fold depletion (n = 1). Values <1 were set to 1 for clarity.

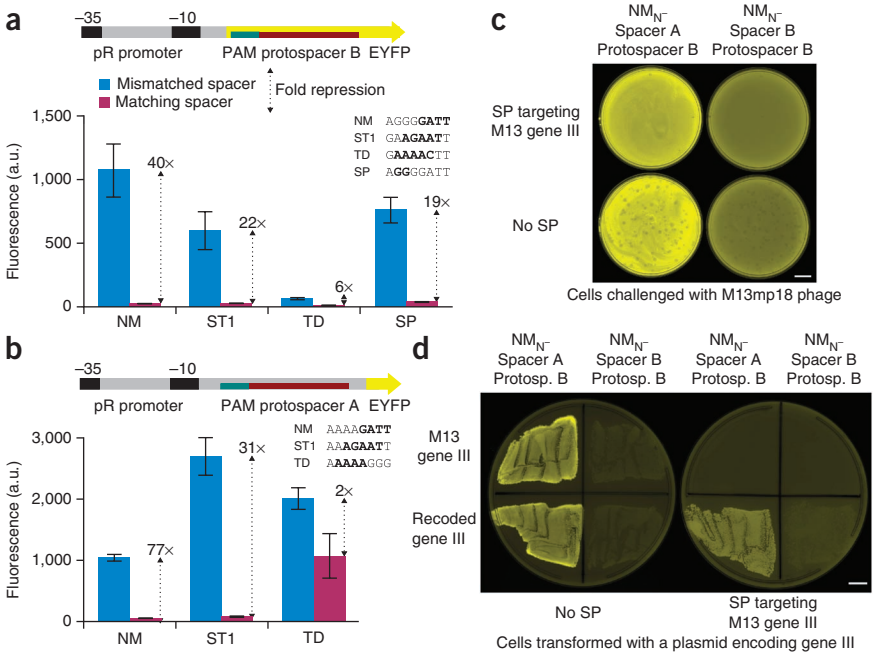

Figure 4 | Simultaneous transcriptional repression and nuclease activity in bacteria. (**a**) Reporter plasmids for quantification of Cas9 repression contained protospacer B and a suitable PAM in the nontemplate strand after the EYFP start codon. Normalized cellular fluorescence is shown for mismatched and matched spacer-protospacer pairs. a.u., arbitrary units. PAMs for each Cas9 are listed. (**b**) Cas9 ortholog repression was verified with a second reporter plasmid containing protospacer A and a PAM in the nontemplate strand within the 5′ UTR. Error bars in **a** and **b** represent the s.d. of eight independently isolated cultures. (**c**) Cells containing the plasmids used for NM-mediated repression (**Fig. 3a**) were transformed with a compatible plasmid encoding SP, its tracrRNA and a five-spacer CRISPR locus designed to cleave filamentous phage gene III at multiple sites and were then challenged with M13mp18. No plaques were visible on bacteria containing the phage defense plasmid, and fluorescence was reduced for cells bearing a matching NM spacer and protospacer relative to the mismatched case. (**d**) Cells were transformed with a compatible plasmid encoding carbenicillin resistance and either wild-type gene III or a recoded version lacking protospacers and were then plated. The combination of the phage defense plasmid and the plasmid encoding wild-type gene III yielded no colonies, whereas all other combinations produced lawns. Fluorescence was reduced for cells bearing a matching NM spacer and protospacer relative to the mismatched case. Scale bars, 10 mm.

our orthogonal Cas9 proteins to mediate multiple independent activities within a single cell.

Genome editing in human cells

We next sought to apply these Cas9 variants to engineer human cells. We constructed sgRNAs from the corresponding crRNAs and tracrRNAs for NM and ST1, the two smaller and more predictably active Cas9 orthologs, by examining complementary regions between crRNA and tracrRNA[29] (**Supplementary Fig. 2**) and fusing the two sequences via a stem-loop at various fusion junctions analogous to those of the sgRNAs created for SP (**Supplementary Fig. 3**). When the existing sequence was liable to cause problems for our expression system (for example, due to multiple successive uracils causing Pol III termination), we generated multiple single-base mutants. The complete 3′ tracrRNA sequence was always included, as truncations are known to be detrimental[8].

We assayed all sgRNAs for activity along with their corresponding Cas9 protein using our previously described homologous recombination assay in 293 cells[8]. Briefly, a genomically integrated nonfluorescent GFP reporter line was constructed for each Cas9 protein in which the GFP coding sequence was interrupted by an insert encoding a stop codon and protospacer sequence with functional PAM. Reporter lines were transfected with expression vectors encoding a Cas9 protein and corresponding sgRNA along with a repair donor capable of restoring fluorescence upon nuclease-induced homologous recombination (**Fig. 5a**). Notably, we observed that full-length crRNA-tracrRNA fusions were active in all instances and therefore represented a reliable method of testing novel Cas9 ortholog activity in eukaryotic cells (**Supplementary Figs. 3–5**). Some but not all truncated versions were equally as active. We selected highly active sgRNA for both NM and ST1 to use in future experiments (**Supplementary Figs. 3–5**).

Cas9 orthogonality in mammalian cells

To verify that none of the three proteins could be guided by the sgRNAs of the others in human cells, we employed the same homologous recombination assay to measure the comparative efficiency of SP, NM and ST1 in combination with each of the three sgRNAs. Of note, NM and ST1 induced genome editing at

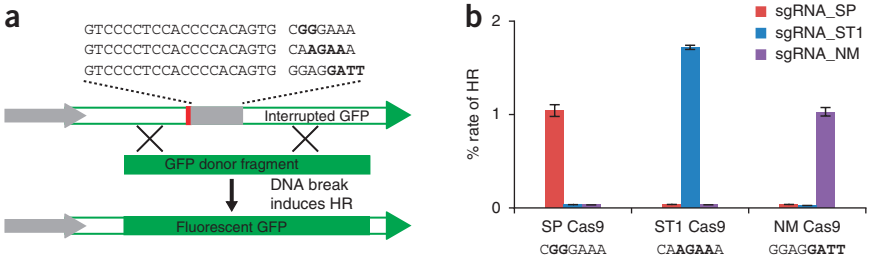

Figure 5 | Cas9-mediated gene editing in human cells. (**a**) A homologous recombination (HR) assay was used to quantify gene-editing efficiency. Cas9-mediated double-strand breaks within the protospacer stimulated repair of an interrupted GFP cassette using the donor template, yielding cells with intact GFP. Three different templates were used to provide the correct PAM for each Cas9. Fluorescent cells were quantified by flow cytometry. (**b**) HR efficiencies for the indicated enzymes and sgRNAs. Substrate PAMs are displayed below each Cas9. Data represent mean ± s.e.m. (n = 3).

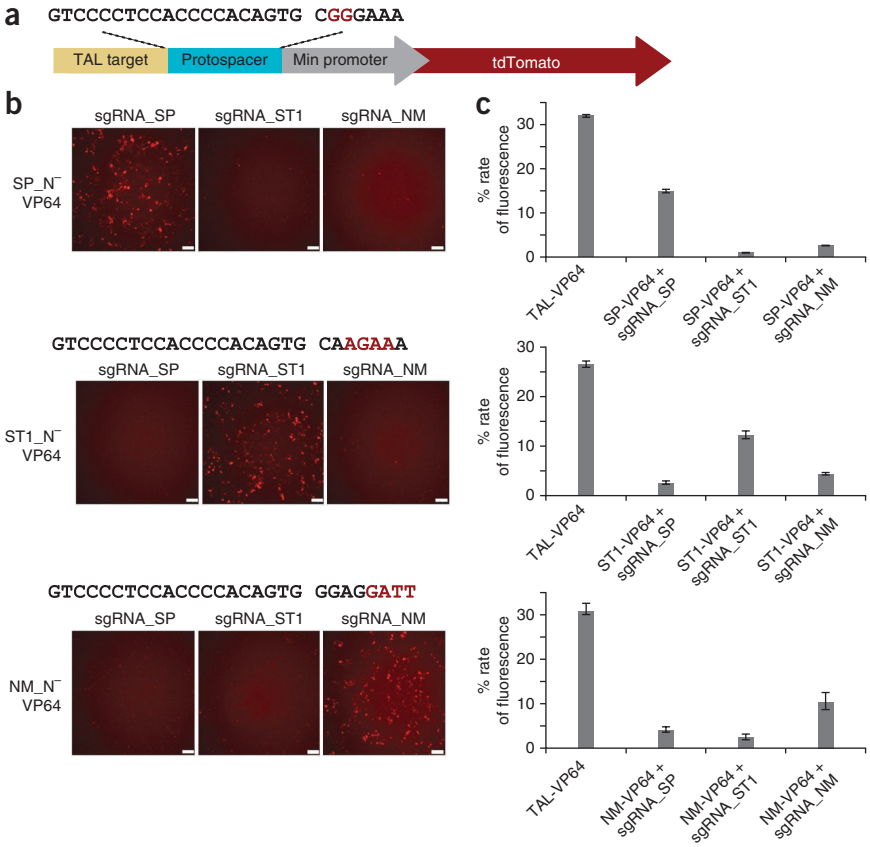

Figure 6 | Transcriptional activation in human cells. (**a**) Reporter constructs for transcriptional activation featured a minimal (Min) promoter driving tdTomato. Nuclease-null (N⁻) Cas9-VP64 fusion proteins binding to upstream protospacers resulted in transcriptional activation and enhanced fluorescence. (**b**) Cells were transfected with all combinations of Cas9 activators and sgRNAs, and tdTomato fluorescence was visualized by microscopy. Scale bars, 100 µm. (**c**) Activation was quantified by flow cytometry along with a TAL (transcription activator–like)-VP64 activator targeting an upstream sequence for comparison. Data represent mean ± s.e.m. (n = 3).

levels comparable to that of SP (**Fig. 5b**). Corroborating our findings with crRNAs in bacteria, our results show that all three Cas9 proteins are fully orthogonal to one another; as such, they are capable of targeting distinct and nonoverlapping sets of sequences within the same cell (**Fig. 5b** and **Supplementary Fig. 6**).

To disentangle the contributions of sgRNA and PAM to orthogonal targeting, we tested a variety of downstream PAM sequences with SP and ST1 and their respective sgRNAs. Certain PAMs were acceptable to both SP and ST1, enabling both enzymes to target the exact same sequence, but cutting occurred only when each enzyme was paired with its corresponding sgRNA. These results highlight the importance of both sgRNA and PAM for Cas9 activity but also emphasize that the specific affinity of each Cas9 for its corresponding sgRNA is sufficient for orthogonality (**Supplementary Fig. 7**).

Transcriptional activation in human cells

We next investigated the ability of NM and ST1 to mediate transcriptional activation in human cells. Nuclease-null NM and ST1 genes were fused to the VP64 activator domain at their C termini to yield putative RNA-guided activators modeled after our SP activator[19]. Reporter constructs for activation consisted of a protospacer with an appropriate PAM inserted upstream of the tdTomato coding region. Vectors expressing an RNA-guided transcriptional activator, an sgRNA and an appropriate reporter were cotransfected, and transcriptional activation was measured by FACS (**Fig. 6**). In each case, we observed robust transcriptional activation by all three Cas9 variants, similar to that of a corresponding transcription activator–like (TAL)-VP64 activator (**Fig. 6**). Each Cas9 activator stimulated transcription only when paired with its corresponding sgRNA, a result confirming orthogonal genome regulation by the three Cas9 proteins.

DISCUSSION

By experimentally characterizing and demonstrating orthogonality among multiple Cas9 proteins in bacteria and human cells, we have substantially expanded the repertoire of orthogonal RNA-guided DNA-binding elements and constructed a pipeline for characterizing additional examples. Together these proteins constitute the basis of a platform enabling simultaneous transcriptional regulation, labeling and gene editing within individual cells.

Our results illustrate the remarkable diversity of proteins within a single family of CRISPR systems. Though clearly related, the Cas9 proteins from *S. pyogenes*, *N. meningitidis*, *S. thermophilus* and *T. denticola* range from 3.25 to 4.6 kbp in length and recognize completely different PAM sequences. These findings are in keeping with the strongly diversifying selective pressures facing defense systems engaged in molecular 'arms races'[30] and suggest that many other Cas9 proteins may be equally orthogonal.

Using two distinct protospacers for comprehensive PAM characterization allowed us a glimpse of the complexities governing protospacer and PAM recognition. Differential protospacer cleavage efficiencies exhibited a consistent trend across diverse Cas9 proteins, although the magnitude of the disparity varied considerably among orthologs. This pattern suggests that sequence-dependent differences in D-loop formation or stabilization determine the basal targeting efficiency for each protospacer but also that additional Cas9 or repeat-dependent factors play a role. Similarly, numerous factors preclude efforts to describe PAM recognition with a single sequence motif. Individual bases adjacent to the primary PAM recognition determinants can combine to dramatically decrease overall affinity. Indeed, certain PAMs appear to interact nonlinearly with the spacer or protospacer to determine the overall activity. Moreover, different affinity levels may be required for distinct activities across disparate cell types. Finally, we observed that our experimentally identified PAMs required fewer bases than did those inferred from bioinformatic analyses, which suggests that spacer acquisition requirements differ from those for effector cleavage.

This difference was most notable for the Cas9 protein from *N. meningitidis*, which had fewer PAM requirements when paired with our spacers than either its bioinformatic prediction or the currently popular Cas9 from *S. pyogenes*, as well as considerably fewer than either ST1 or TD. It would be interesting to determine whether the total protospacer-PAM specificity of these four proteins is related to organismal genome size—a relationship that could point toward more specific Cas9 orthologs. More immediately, the characterization of NM considerably expands the number of sequences that can be readily targeted with a Cas9 protein. At 3.25 kbp in length, the NM gene is also 850 bp smaller than the SP gene; both the NM and ST1 genes are small enough to fit into standard viral vectors for *in vivo* delivery. NM may represent a more suitable starting point for directed evolution efforts designed to alter PAM recognition or specificity. We expect future experiments aimed at characterizing additional Cas9 orthologs to further improve our mechanistic understanding and expand our engineering capabilities.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Constructs are available on Addgene (see **Supplementary Table 1**): 48645–48679.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank P.B. Stranges for protein alignments and W.L. Chew for helpful discussions. This work was supported by US National Institutes of Health NHGRI grant P50 HG005550, US Department of Energy grant DE-FG02-02ER63445 and the Wyss Institute for Biologically Inspired Engineering.

AUTHOR CONTRIBUTIONS

K.M.E. and P.M. conceived of the study; K.M.E. and P.M. designed the experiments; K.M.E., J.L.B. and S.J.Y. performed experiments in *E. coli*; J.L.B. wrote analysis software; P.M. and M.M. performed experiments in human cells; K.M.E. and P.M. analyzed results; and K.M.E. and P.M. wrote the manuscript with input from G.M.C.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. *Annu. Rev. Genet.* **45**, 273–297 (2011).
2. Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
3. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. USA* **109**, E2579–E2586 (2012).
4. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
5. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
6. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
7. Ding, Q. *et al.* Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell* **12**, 393–394 (2013).
8. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
9. Wang, H. *et al.* One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
10. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
11. Boch, J. *et al.* Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509–1512 (2009).
12. Gaj, T., Gersbach, C.A. & Barbas, C.F. III. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* **31**, 397–405 (2013).
13. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.* **27**, 851–857 (2009).
14. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA* **93**, 1156–1160 (1996).
15. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
16. Porteus, M.H. & Carroll, D. Gene targeting using zinc finger nucleases. *Nat. Biotechnol.* **23**, 967–973 (2005).
17. Urnov, F.D. *et al.* Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
18. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
19. Gilbert, L.A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
20. Mali, P. *et al.* CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat. Biotechnol.* **31**, 833–838 (2013).
21. Maeder, M.L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nat. Methods* doi:10.1038/nmeth.2598 (25 July 2013).
22. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat. Methods* doi:10.1038/nmeth.2600 (25 July 2013).
23. Podgornaia, A.I. & Laub, M.T. Determinants of specificity in two-component signal transduction. *Curr. Opin. Microbiol.* **16**, 156–162 (2013).
24. Purnick, P.E. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10**, 410–422 (2009).
25. Horvath, P. *et al.* Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1401–1412 (2008).
26. Zhang, Y. *et al.* Processing-independent CRISPR RNAs limit natural transformation in *Neisseria meningitidis*. *Mol. Cell* **50**, 488–503 (2013).
27. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
28. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
29. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
30. Bondy-Denomy, J., Pawluk, A., Maxwell, K.L. & Davidson, A.R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).

ONLINE METHODS

Vector and strain construction. Cas9 sequences from *S. thermophilus*, *N. meningitidis*, and *T. denticola* were obtained from NCBI and human-codon optimized using JCAT (http://www.jcat.de/)[31] and modified to facilitate DNA synthesis and expression in *E. coli*. 500-bp gBlocks (Integrated DNA Technologies) were joined by hierarchical overlap PCR and isothermal assembly[27]. The resulting full-length products were subcloned into bacterial and human expression vectors. Nuclease-null Cas9 cassettes (NM: D16A D587A H588A N611A, SP: D10A D839A H840A N863A, ST1: D9A D598A H599A N622A, TD: D13A D878A H879A N902A) were constructed from these templates by standard methods.

Bacterial plasmids. Cas9 was expressed in bacteria from a cloDF13-aadA plasmid backbone using the medium-strength *proC* constitutive promoter. All tracrRNA cassettes (**Supplementary Note 2**), including promoters and terminators from the native bacterial loci, were synthesized as gBlocks and inserted downstream of the Cas9 coding sequence for each vector for robust tracrRNA production. When the tracrRNA cassette was expected to additionally contain a promoter in the opposite orientation, the lambda t1 terminator was inserted to prevent interference with *cas9* transcription. Bacterial targeting plasmids were based on a p15A-cat backbone with the strong J23100 promoter followed by one of two 20-bp spacer sequences (**Fig. 2d**) previously determined to function using SP. Substrate plasmids for orthogonality testing in bacteria were identical to library plasmids (see below) but with the following PAMs: GAAGGGTT (NM), GGGAGGTT (SP), GAAGAATT (ST1), AAAAAGGG (TD). Spacer sequences were immediately followed by one of the three 36-bp repeat sequences depicted in **Figure 1a**. EYFP reporter vectors were based on a pSC101-kan backbone with the pR promoter driving GFP and the T7 g10 RBS preceding the EYFP coding sequence. Two types were created: one with protospacer B + PAM inserted into the nontemplate strand just after the start codon of EYFP, and one with protospacer A + PAM inserted into the nontemplate strand in the 5′ UTR. PAMs are as listed in **Figure 4a**,**b**. The plasmid conferring immunity to filamentous phages via SP features a colE1-erm backbone, the SP cas9 gene, and tracrRNA exactly as in the standard cloDF13-aadA plasmids and the J23100 promoter driving a CRISPR locus targeting five sites within M13 gene III. Transformed plasmids carry the *bla* gene for carbenicillin resistance and either wild-type gene III or a recoded gene III. The CRISPR locus and recoded gene III were synthesized by Genewiz. Sequences of *cas9* genes used in bacterial experiments are provided (**Supplementary Note 3**). All vectors and complete sequences are available through Addgene (**Supplementary Table 1**).

Mammalian vectors. Mammalian Cas9 expression vectors were based on pcDNA3.3-TOPO with C-terminal SV40 NLSs. sgRNAs for each Cas9 were designed by aligning crRNA repeats with tracrRNAs and fusing the 5′ crRNA repeat to the 3′ tracrRNA so as to leave a stable stem for Cas9 interaction (**Supplementary Table 2**)[29]. sgRNA expression constructs were generated by cloning the U6-sgRNA expressing fragments synthesized as gBlocks into the pCR-BluntII-TOPO vector backbone. Spacers were identical to those used in previous work[8]. Lentivectors for the broken-GFP HR reporter assay were modified from those previously described to include appropriate PAM sequences for each Cas9 and used to establish the stable GFP reporter lines.

RNA-guided transcriptional activators consisted of nuclease-null Cas9 proteins fused to the VP64 activator and corresponding reporter constructs bearing a tdTomato driven by a minimal promoter were constructed as previously described[20]. Sequences of *cas9* genes used in mammalian experiments are provided (**Supplementary Note 4**). All vectors and complete sequences are available through Addgene (**Supplementary Table 1**).

Library construction and transformation. Protospacer libraries were constructed by amplifying the pZE21 vector (ExpressSys) using primers (IDT) encoding one of the two protospacer sequences followed by eight random bases and assembled by standard isothermal methods[27]. Library assemblies were initially transformed into NEBTurbo cells (New England Biolabs), yielding $>1 \times 10^8$ clones per library according to dilution plating, and purified by Midiprep (Qiagen). Electrocompetent NEBTurbo cells containing a Cas9 expression plasmid (DS-NMcas, DS-ST1cas, or DS-TDcas) and a targeting plasmid (PM-NM!sp1, PM-NM!sp2, PM-ST1!sp1, PM-ST1!sp2, PM-TD!sp1, or PM-TD!sp2) were transformed with 200 ng of each library and recovered for 2 h at 37 °C before dilution with medium containing spectinomycin (50 μg/mL), chloramphenicol (30 μg/mL) and kanamycin (50 μg/mL). Serial dilutions were plated to estimate post-transformation library size. All libraries exceeded $\sim 1 \times 10^7$ clones, indicative of complete coverage of the 65,536 random PAM sequences.

High-throughput sequencing. Library DNA was harvested by spin columns (Qiagen) after 12 h of antibiotic selection. Intact PAMs were amplified with barcoded primers (**Supplementary Table 3**) and sequences obtained from overlapping 25-bp paired-end reads on an Illumina MiSeq. MiSeq yielded 18,411,704 total reads or 9,205,852 paired-end reads with an average quality score >34 for each library. Paired-end reads were merged and filtered for perfect alignment to each other, their protospacer, and the plasmid backbone. The remaining 7,652,454 merged filtered reads were trimmed to remove plasmid backbone and protospacer sequences and then used to generate position weight matrices for each PAM library. Each library combination received at least 450,000 high-quality reads.

Sequence processing. To calculate the fold depletion for each candidate PAM, we employed two scripts to filter the data. patternProp (**Supplementary Software 1**, usage: python patternProp.py [PAM] file.fastq) returns the number and fraction of reads matching each one-base derivative of the indicated PAM. One-base derivatives are defined as the set of all sequences in which one additional base that was not specified in the parent (i.e., N) is set to A, C, G, or T. patternProp3 (**Supplementary Software 2**) returns the fraction of reads matching each one-base derivative relative to the total number of reads for the library. Spreadsheets detailing depletion ratios for each calculated PAM (**Supplementary Data**) were used to identify the minimal fold depletion among all one-base derivatives and thereby classify PAMs. Total reads were tabulated and barcode clusters examined (**Supplementary Data**) to verify library sizes.

doi:10.1038/nmeth.2681

Repression and orthogonality assays in bacteria. Cas9-mediated repression was assayed by transforming the NM expression plasmid and the EYFP reporter plasmid with each of the two corresponding targeting plasmids. Colonies with matching or mismatched spacer and protospacer were picked and grown in 96-well plates. Fluorescence at 495–528 nm and absorbance at 600 nm were measured using a Synergy Neo microplate reader (BioTek).

Orthogonality tests were performed by preparing electrocompetent NEBTurbo cells bearing all combinations of Cas9 and targeting plasmids and transforming them with matched or mismatched substrate plasmids bearing appropriate PAMs for each Cas9. Sufficient cells and dilutions were plated to ensure that at least some colonies appeared even for correct Cas9 + targeting + matching protospacer combinations, which typically arise owing to mutational inactivation of the Cas9 or the crRNA. Colonies were counted and fold depletion calculated for each.

For the simultaneous nuclease and repression assays, cells were first rendered electrocompetent, transformed with the SP phage defense plasmid, and plated with erythromycin, kanamycin, chloramphenicol, and spectinomycin. Plaque assays were performed by mixing dilutions of M13mp18 phage (New England Biolabs) with 75 µL of cells (NEBTurbo, containing the F plasmid), combining with 1 mL soft agar, and plating onto 60-mm LB plates with 50 µg/mL IPTG and 200 µg/mL X-Gal. For the plasmid transformation assay, cells were rendered electrocompetent by standard methods, transformed with plasmids bearing wild-type or recoded gene III, and plated with carbenicillin, erythromycin, kanamycin, chloramphenicol, and spectinomycin. Plaque assays were imaged using a Typhoon FLA 9000 (GE), and the contrast adjusted by setting the maximum saturation to 1.0% using ImageJ.

Cell culture and transfections. HEK 293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Invitrogen) with high glucose supplemented with 10% FBS (FBS, Invitrogen), penicillin/streptomycin (pen/strep, Invitrogen), and non-essential amino acids (NEAA, Invitrogen). Cells were maintained at 37 °C and 5% CO_2 in a humidified incubator.

Transfections involving nuclease assays were as follows: 0.4×10^6 cells were transfected with 2 µg Cas9 plasmid, 2 µg gRNA and/or 2 µg DNA donor plasmid using Lipofectamine 2000 per the manufacturer's protocols. Cells were harvested 3 d after transfection and either analyzed by FACS, or, for direct assay of genomic cuts, the genomic DNA of ~1×10^6 cells was extracted using DNAeasy kit (Qiagen).

For transfections involving transcriptional activation assays: 0.4×10^6 cells were transfected with 2 µg $Cas9_{N-}$VP64 plasmid, 2 µg gRNA and/or 0.25 µg of reporter construct. Cells were harvested 24–48 h post transfection and assayed using FACS or immunofluorescence methods, or their total RNA was extracted, and these were subsequently analyzed by RT-PCR.

No samples were excluded from any experiments performed in this study.

31. Grote, A. *et al.* JCat: a novel tool to adapt codon usage of a target gene to its potential expression host. *Nucleic Acids Res.* **33**, W526–W531 (2005).