nature biotechnology

RNA-guided editing of bacterial genomes using CRISPR-Cas systems

Wenyan Jiang[1,4], David Bikard[1,4], David Cox[2,3], Feng Zhang[2,3] & Luciano A Marraffini[1]

Here we use the clustered, regularly interspaced, short palindromic repeats (CRISPR)–associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli*. The approach relies on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-selection systems. We reprogram dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. Simultaneous use of two crRNAs enables multiplex mutagenesis. In *S. pneumoniae*, nearly 100% of cells that were recovered using our approach contained the desired mutation, and in *E. coli*, 65% that were recovered contained the mutation, when the approach was used in combination with recombineering. We exhaustively analyze dual-RNA:Cas9 target requirements to define the range of targetable sequences and show strategies for editing sites that do not meet these requirements, suggesting the versatility of this technique for bacterial genome engineering.

The understanding of gene function depends on the possibility of altering DNA sequences within the cell in a controlled fashion. Site-specific mutagenesis in eukaryotes is achieved through the use of sequence-specific nucleases that promote homologous recombination of a template DNA containing the mutation of interest. Zinc finger nucleases (ZFNs)[1], transcription activator–like effector nucleases (TALENs)[2] and homing meganucleases[3] can be programmed to cleave genomes in specific locations, but these approaches require engineering of new enzymes for each target sequence. In prokaryotic organisms, mutagenesis methods either introduce a selection marker in the edited locus or require a two-step process that includes a counter-selection system[4,5]. More recently, phage recombination proteins have been used for recombineering, a technique that promotes homologous recombination of linear DNA or oligonucleotides. However, because there is no selection of mutations, recombineering efficiency can be relatively low (0.1–10% for point mutations down to 10^{-5}–10^{-6} for larger modifications)[6], in many cases requiring the screening of a large number of colonies. Therefore new technologies that are affordable, easy to use and efficient are still needed for the genetic engineering of both eukaryotic and prokaryotic organisms.

Recent work on the CRISPR adaptive immune system of prokaryotes has led to the identification of nucleases whose sequence specificity is programmed by small RNAs[7]. CRISPR loci are composed of a series of repeats separated by 'spacer' sequences that match the genomes of bacteriophages and other mobile genetic elements[8–11]. The repeat-spacer array is transcribed as a long precursor and processed within repeat sequences to generate small crRNA that specify the target sequences (also known as protospacers) cleaved by CRISPR systems[12–16]. Essential for cleavage is the presence of a sequence motif immediately downstream of the target region, known as the protospacer-adjacent motif (PAM)[7,17,18]. CRISPR-associated (*cas*) genes usually flank the repeat-spacer array and encode the enzymatic machinery responsible for crRNA biogenesis and targeting[19]. Cas9 is a double-stranded (ds)DNA endonuclease that uses a crRNA guide to specify the site of cleavage[7,18]. Loading of the crRNA guide onto Cas9 occurs during the processing of the crRNA precursor and requires a small RNA antisense to the precursor (the tracrRNA) and RNAse III (ref. 14). In contrast to genome editing with ZFNs or TALENs, changing the target specificity of the RNA-protein complex comprised of tracrRNA, crRNA and Cas9 (dual-RNA:Cas9) does not require protein engineering but only the design of the short crRNA guide[20–22].

We recently showed in *S. pneumoniae* that the introduction of a CRISPR-Cas system targeting a chromosomal locus leads to the killing of the transformed cells[23]. We observed that occasional survivors contained mutations in the target region, suggesting that dual-RNA:Cas9 endonuclease activity against endogenous targets could be used for genome editing. Here we show that marker-less mutations can be introduced through the transformation of a template DNA fragment that will recombine in the genome and eliminate recognition of the target by the endonuclease. Directing the specificity of dual-RNA:Cas9 with several different crRNAs allows for the introduction of multiple mutations at the same time. We also characterize in detail the sequence requirements for dual-RNA:Cas9 targeting and show that the approach can be combined with recombineering for genome editing in *E. coli*.

RESULTS

Genome editing by dual-RNA:Cas9 cleavage of a genomic target

S. pneumoniae strain crR6 contains a dual-RNA:Cas9-based CRISPR system that cleaves a target sequence present in the bacteriophage ϕ8232.5. This target is integrated into the *srtA* chromosomal locus of a second *S. pneumoniae* strain $R6^{8232.5}$. An altered target sequence containing a mutation in the PAM region is integrated into the *srtA*

[1]Laboratory of Bacteriology, The Rockefeller University, New York, New York, USA. [2]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]McGovern Institute for Brain Research, Department of Brain and Cognitive Science and Department of Biological Engineering, MIT, Cambridge, Massachusetts, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to D.B. (dbikard@rockefeller.edu) or L.A.M. (marraffini@rockefeller.edu).

Received 29 November 2012; accepted 16 January 2013; published online 29 January 2013; doi:10.1038/nbt.2508

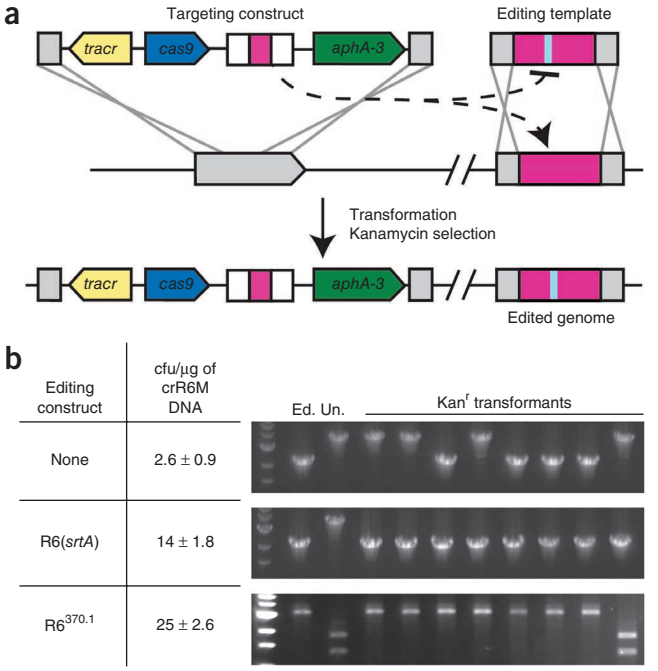

Figure 1 dual-RNA:Cas9 nuclease activity against endogenous targets can be exploited for genome editing. (**a**) Concept of genome editing using the CRISPR-Cas system. The CRISPR targeting construct directs cleavage of a chromosomal locus and is co-transformed with an editing template that recombines with the target to prevent cleavage. Kanamycin-resistant transformants that survive CRISPR attack contain modifications introduced by the editing template. *tracr*, trans-activating CRISPR RNA; *aphA-3*, kanamycin resistance gene. (**b**) Transformation of crR6M DNA in $R6^{8232.5}$ cells with no editing template, the R6 wild-type *srtA* or the $R6^{370.1}$ editing templates. Transformation efficiency is calculated as colony forming units (cfu) per µg of crR6M DNA; the mean values with s.d. from at least three independent experiments are shown. PCR analysis is performed on eight clones in each transformation. Un., the unedited *srtA* locus of strain $R6^{8232.5}$; Ed., the editing template. $R6^{8232.5}$ and $R6^{370.1}$ targets are distinguished by restriction with EaeI.

locus of a third *S. pneumoniae* strain, $R6^{370.1}$, abolishing cleavage by dual-RNA:Cas9 (ref. 23)(**Supplementary Fig. 1a**). We transformed $R6^{8232.5}$ and $R6^{370.1}$ cells with genomic DNA from crR6 cells, expecting that successful transformation of $R6^{8232.5}$ cells (but not $R6^{370.1}$ cells) should lead to cleavage of the target locus and cell death. Contrary to this expectation, we isolated $R6^{8232.5}$ transformants, albeit with approximately tenfold less efficiency than $R6^{370.1}$ transformants (**Supplementary Fig. 1b**). Genetic analysis of eight $R6^{8232.5}$ transformants (**Supplementary Fig. 1b**) revealed that the great majority are the product of a double recombination event that eliminates the toxicity of dual-RNA:Cas9 targeting by replacing the ϕ8232.5 target with the crR6 genome's wild-type *srtA* locus, which does not contain the protospacer required for dual-RNA:Cas9 recognition. These results are proof that the concurrent introduction of a CRISPR-Cas system targeting a genomic locus (the targeting construct), together with a template for recombination into the targeted locus (the editing template), can lead to targeted genome editing (**Fig. 1a**).

To create a simplified system for genome editing, we modified the CRISPR locus in strain crR6 by deleting *cas1*, *cas2* and *csn2*, genes which have been shown to be dispensable for dual-RNA:Cas9 targeting[14,24], yielding strain crR6M (**Supplementary Fig. 1a**). This strain retains the properties of crR6 (**Supplementary Fig. 1b**). To increase the efficiency of editing and demonstrate that a template DNA of choice can be used to control the mutation introduced, we co-transformed $R6^{8232.5}$ cells with PCR products of the wild-type *srtA* gene or the mutant $R6^{370.1}$ target, either of which should be resistant to cleavage by dual-RNA:Cas9. This resulted in a five- to tenfold increase of the frequency of transformation compared with genomic crR6 DNA alone (**Fig. 1b**). The efficiency of editing was also substantially increased, with 8/8 transformants tested containing a wild-type *srtA* copy and 7/8 containing the PAM mutation present in the $R6^{370.1}$ target (**Fig. 1b** and **Supplementary Fig. 2a**). Taken together, these results show the potential of genome editing assisted by a CRISPR-Cas system.

Analysis of dual-RNA:Cas9 target requirements

To introduce specific changes in the genome, one must use an editing template carrying mutations that abolish dual-RNA:Cas9-mediated cleavage, thereby preventing cell death. This is easy to achieve when the deletion of the target or its replacement by another sequence (gene insertion) is sought. When the goal is to produce gene fusions or to generate single-nucleotide mutations, the prevention of cleavage by dual-RNA:Cas9 is possible only by introducing mutations in the editing template that alter either the PAM or the protospacer sequences. To determine the constraints imposed by these sequences, we performed a thorough analysis of PAM and protospacer mutations that abrogate dual-RNA:Cas9 targeting.

Previous studies proposed that *S. pyogenes* dual-RNA:Cas9 requires an NGG PAM immediately downstream of the protospacer[7,14,18]. However, because only a very limited number of PAM-inactivating mutations have been described so far[7,14,18,23], we conducted a systematic analysis to find all 5-nucleotide sequences following the protospacer that eliminate dual-RNA:Cas9 cleavage. We used randomized oligonucleotides to generate all possible 1,024 PAM sequences in a heterogeneous PCR product that was used to transform crR6 or R6 cells. Constructs carrying functional PAMs are expected to be recognized and destroyed in crR6 but not R6 cells (**Fig. 2a**). More than 2×10^5 colonies were pooled to extract DNA for use as a template for the co-amplification of all targets. PCR products were deep sequenced and found to contain all 1,024 sequences, with coverage ranging from 5 to 42,472 reads (**Supplementary Data**). The functionality of each PAM was estimated by the relative proportion of its reads in the crR6 sample over the R6 sample. Analysis of the first three bases of the PAM clearly shows that the NGG pattern is underrepresented in crR6 transformants (**Fig. 2b**). Furthermore, the next two bases have no detectable effect on the NGG PAM (**Supplementary Discussion**), demonstrating that the NGGNN sequence is sufficient to license dual-RNA:Cas9 activity. Partial targeting was observed for NAG PAM sequences (**Fig. 2b**). Also the NNGGN pattern partially inactivates targeting (**Supplementary Table 1**), indicating that the NGG motif can still be recognized by dual-RNA:Cas9 with reduced efficiency when shifted by 1 bp. These data shed light onto the molecular mechanism of dual-RNA:Cas9 target recognition, and they reveal that NGG (or CCN on the complementary strand) sequences are sufficient for targeting and that NGG to NAG or NNGGN mutations in the editing template should be avoided. Owing to the high frequency of these trinucleotide sequences (once every 8 bp), this means that almost any position of the genome can be edited. Indeed, we tested ten randomly chosen targets carrying various PAMs and all were found to be functional (**Supplementary Fig. 3**).

Another way to disrupt dual-RNA:Cas9-mediated cleavage is to introduce mutations in the protospacer region of the editing template. It is known that point mutations within the 'seed sequence' (the 8–10 protospacer nucleotides immediately adjacent to the PAM)

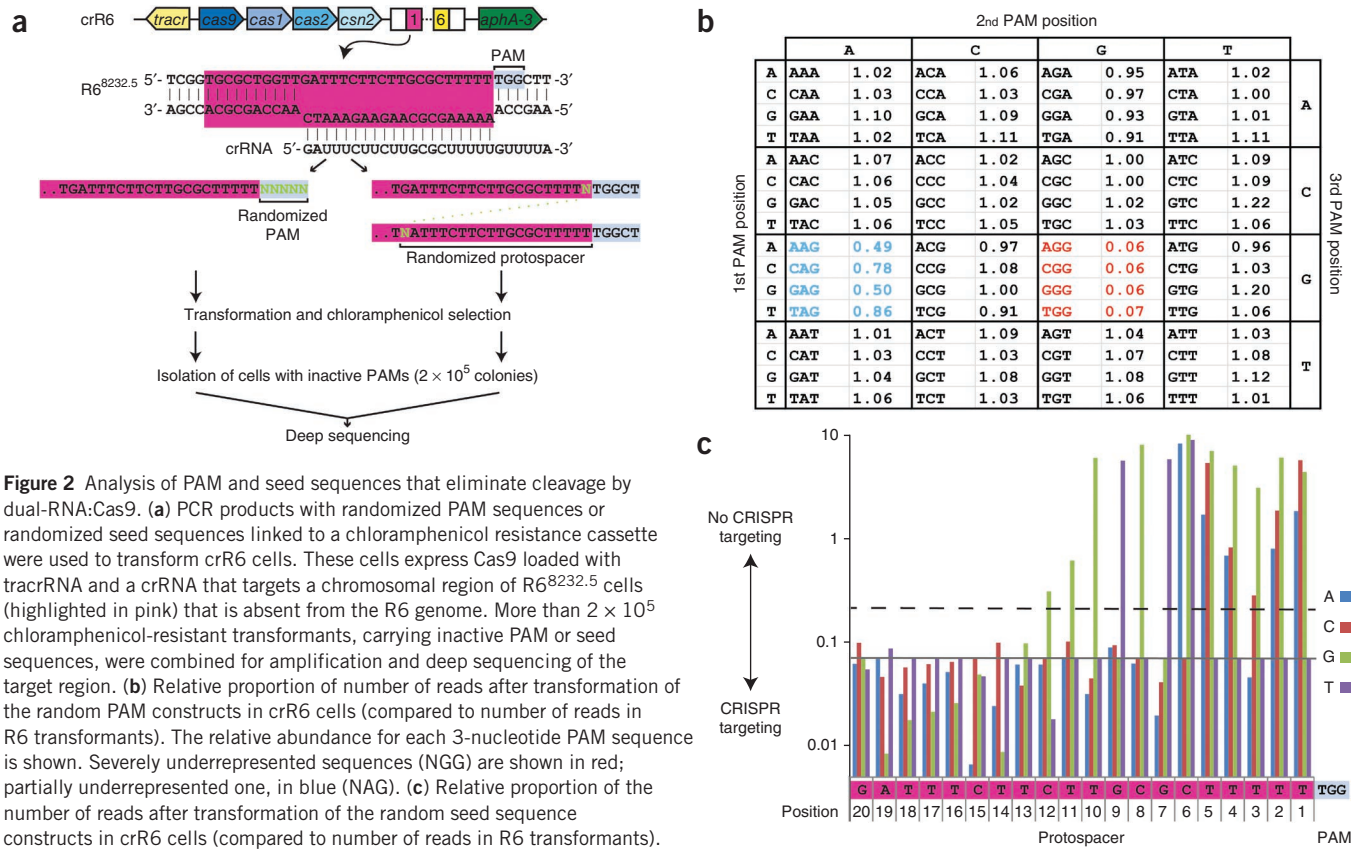

Figure 2 Analysis of PAM and seed sequences that eliminate cleavage by dual-RNA:Cas9. (**a**) PCR products with randomized PAM sequences or randomized seed sequences linked to a chloramphenicol resistance cassette were used to transform crR6 cells. These cells express Cas9 loaded with tracrRNA and a crRNA that targets a chromosomal region of R6$^{8232.5}$ cells (highlighted in pink) that is absent from the R6 genome. More than 2×10^5 chloramphenicol-resistant transformants, carrying inactive PAM or seed sequences, were combined for amplification and deep sequencing of the target region. (**b**) Relative proportion of number of reads after transformation of the random PAM constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance for each 3-nucleotide PAM sequence is shown. Severely underrepresented sequences (NGG) are shown in red; partially underrepresented one, in blue (NAG). (**c**) Relative proportion of the number of reads after transformation of the random seed sequence constructs in crR6 cells (compared to number of reads in R6 transformants). The relative abundance of each nucleotide for each position of the first 20 nucleotides of the protospacer sequence is shown. High abundance indicates lack of cleavage by dual-RNA:Cas9, that is, an inactivating mutation. The gray line shows the level of the WT sequence. The dashed line represents the level above which a mutation significantly disrupts cleavage (**Supplementary Discussion**).

can abolish cleavage by CRISPR nucleases[7,25,26]. However, the exact length of this region is not known, and it is unclear whether mutations to any nucleotide in the seed can disrupt dual-RNA:Cas9 target recognition. We followed the same deep sequencing approach described above to randomize the entire protospacer sequence involved in base pair contacts with the crRNA and to determine all sequences that disrupt targeting. Each position of the 20 matching nucleotides[14] in the *spc1* target present in R6$^{8232.5}$ cells (**Fig. 1a**) was randomized, and a library containing the resulting sequences was used to transform crR6 and R6 cells (**Fig. 2a**). Consistent with the presence of a seed sequence, only mutations in the 12 nucleotides immediately upstream of the PAM abrogated cleavage (**Fig. 2c**). However, different mutations displayed markedly different effects. The distal (from the PAM) positions of the seed (12 to 7) tolerated most mutations and only one particular base substitution abrogated targeting. In contrast, mutations to any nucleotide in the proximal positions (6 to 1, except 3) prevented cleavage, although at different levels for each particular substitution. At position 3, only two substitutions affected dual-RNA:Cas9 activity and with different strength. We conclude that, although seed sequence mutations can prevent dual-RNA:Cas9 targeting, there are restrictions regarding the nucleotide changes that can be made in each position of the seed. Moreover, these restrictions can most likely vary for different spacer sequences. Therefore we believe that mutations in the PAM sequence, if possible, should be the preferred editing strategy. Alternatively, multiple mutations in the seed sequence could be introduced to prevent dual-RNA:Cas9 nuclease activity.

dual-RNA:Cas9-mediated genome editing in *S. pneumoniae*

To develop a rapid and efficient method for targeted genome editing, we engineered strain crR6Rk, a strain in which spacers can be easily introduced by PCR (**Supplementary Fig. 4**). We decided to edit the β-galactosidase (*bgaA*) gene of *S. pneumoniae*, whose activity can be easily measured[27]. We introduced alanine substitutions of amino acids in the active site of this enzyme: R481A (R→A) and N563A,E564A (NE→AA) mutations. To illustrate different editing strategies, we designed mutations of both the PAM sequence and the protospacer seed. In both cases we used the same targeting construct with a crRNA complementary to a region of the β-galactosidase gene that is adjacent to a TGG PAM sequence (CCA in the complementary strand, **Fig. 3a**). The R→A editing template created a three-nucleotide mismatch on the protospacer seed sequence (CGT to GCA, also introducing a BtgZI restriction site). In the NE→AA editing template we simultaneously introduced a synonymous mutation that creates an inactive PAM (TGG to TTG) along with mutations that are 218 nt downstream of the protospacer region (AAT GAA to GCT GCA, also generating a TseI restriction site). This last editing strategy demonstrates the possibility of using a remote PAM to make mutations in places where a proper target might be hard to choose. For example, although the *S. pneumoniae* R6 genome, which has a 39.7% GC content, contains on average one PAM motif every 12 bp, some PAM motifs are separated by up to 194 bp (**Supplementary Fig. 5**). In addition we designed a Δ*bgaA* in-frame deletion of 6,664 bp. In all three cases, co-transformation of the targeting construct and editing template produced ten times more kanamycin-resistant cells than co-transformation

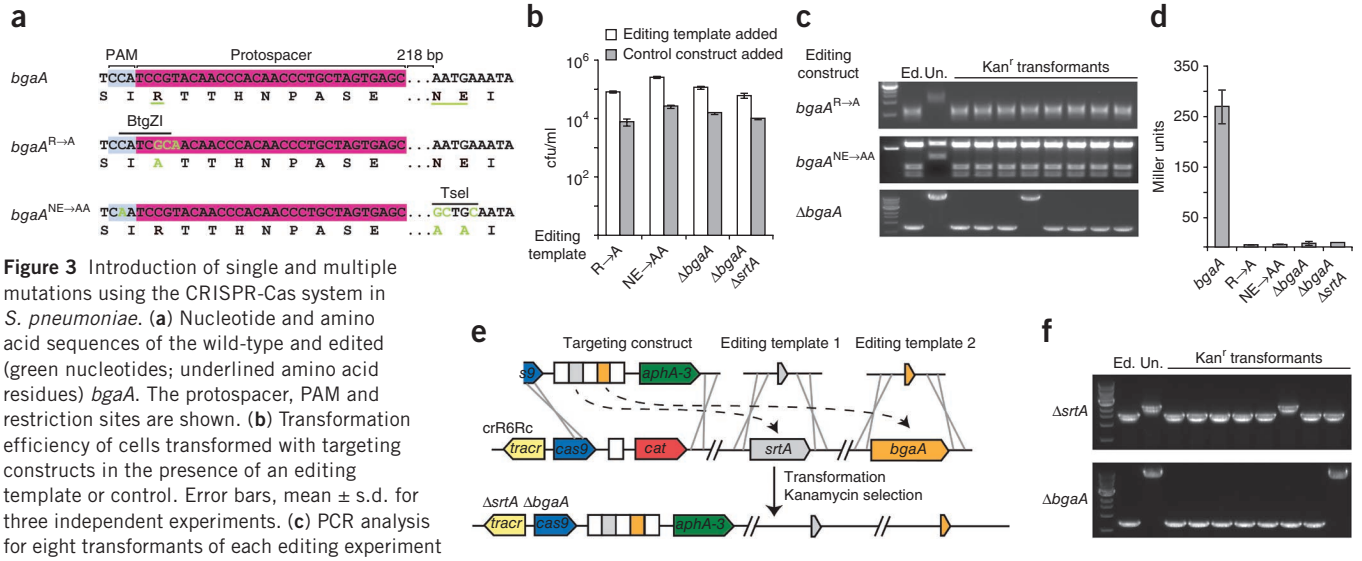

Figure 3 Introduction of single and multiple mutations using the CRISPR-Cas system in *S. pneumoniae*. (**a**) Nucleotide and amino acid sequences of the wild-type and edited (green nucleotides; underlined amino acid residues) *bgaA*. The protospacer, PAM and restriction sites are shown. (**b**) Transformation efficiency of cells transformed with targeting constructs in the presence of an editing template or control. Error bars, mean ± s.d. for three independent experiments. (**c**) PCR analysis for eight transformants of each editing experiment followed by digestion with BtgZI (R→A) and TseI (NE→AA). Deletion of *bgaA* is revealed as a smaller PCR product. (**d**) Miller assay to measure the β-galactosidase activity of WT and edited strains. Error bars, mean ± s.d. for three independent experiments. (**e**) For a single-step, double deletion, the targeting construct contains two spacers (in this case matching *srtA* and *bgaA*) and is co-transformed with two different editing templates. (**f**) PCR analysis for eight transformants to detect deletions in *srtA* and *bgaA* loci. 6/8 transformants contained deletions of both genes.

with a control editing template containing wild-type *bgaA* sequences (**Fig. 3b**). We genotyped 24 transformants (8 for each editing experiment) and found that all but one incorporated the desired change (**Fig. 3c**). DNA sequencing also confirmed not only the presence of the introduced mutations but also the absence of secondary mutations in the target region (**Supplementary Fig. 1b**,**c**). Finally, we measured β-galactosidase activity[27] to confirm that all edited cells displayed the expected phenotype (**Fig. 3d**).

dual-RNA:Cas9-mediated editing can also be used to generate multiple mutations for the study of biological pathways. We decided to illustrate this for the sortase-dependent pathway that anchors surface proteins to the envelope of Gram-positive bacteria[28]. We introduced a sortase deletion by co-transformation of a chloramphenicol-resistant targeting construct and a Δ*srtA* editing template (**Supplementary Fig. 6a**,**b**), followed by a Δ*bgaA* deletion using a kanamycin-resistant targeting construct that replaces the previous one. In *S. pneumoniae*, β-galactosidase is covalently linked to the cell wall by sortase[27]. Therefore, deletion of *srtA* results in the release of the surface protein into the supernatant, whereas the double deletion has no detectable β-galactosidase activity (**Supplementary Fig. 7c**). Such a sequential selection can be repeated as many times as required to generate multiple mutations.

These two mutations can also be introduced at the same time. We designed a targeting construct containing two spacers, one matching *srtA* and the other matching *bgaA*, and co-transformed it with both editing templates at the same time (**Fig. 3e**). Genetic analysis of transformants showed that editing occurred in 6/8 cases (**Fig. 3f**). Notably, the remaining two clones each contained either a Δ*srtA* or a Δ*bgaA* deletion, suggesting the possibility of performing combinatorial mutagenesis using our strategy. Finally, to eliminate the targeting construct sequences, we introduced a plasmid containing the *bgaA* target and a spectinomycin resistance gene along with genomic DNA from the wild-type strain R6. Spectinomycin-resistant transformants that retain the plasmid eliminated the targeting construct sequences (**Supplementary Fig. 7a**,**d**).

Mechanism and efficiency of editing

To understand the mechanisms underlying genome editing with our CRISPR-Cas system, we designed an experiment in which the editing efficiency could be measured independently of cleavage by dual-RNA:Cas9. We integrated the *ermAM* erythromycin resistance gene in the *srtA* locus, and introduced a premature stop codon using dual-RNA:Cas9-mediated editing (**Supplementary Fig. 6**). The resulting strain (JEN53) contains an *ermAM(stop)* allele and is sensitive to erythromycin. This strain can be used to assess the efficiency at which the *ermAM* gene is repaired by measuring the fraction of cells that restore antibiotic resistance with or without the use of cleavage by dual-RNA:Cas9. JEN53 was transformed with an editing template that restores the wild-type allele, together with either a kanamycin-resistant CRISPR construct targeting the *ermAM(stop)* allele (CRISPR::ermAM(stop)) or a control construct without a spacer (CRISPR::Ø) (**Fig. 4a**,**b**). In the absence of kanamycin selection, the fraction of edited colonies is on the order of 10^{-2} (erythromycin-resistant colony-forming units (cfu)/total cfu) (**Fig. 4c**), representing the baseline frequency of recombination without dual-RNA:Cas9-mediated selection against unedited cells. However, if kanamycin selection is applied and the control CRISPR construct is co-transformed, the fraction of edited colonies increases to about 10^{-1} (kanamycin- and erythromycin-resistant cfu/kanamycin-resistant cfu) (**Fig. 4c**). This result shows that selection for the recombination of the CRISPR locus co-selected for recombination in the *ermAM* locus independently of dual-RNA:Cas9 cleavage of the genome, suggesting that a subpopulation of cells is more prone to transformation and/or recombination. Transformation of the CRISPR::ermAM(stop) construct followed by kanamycin selection resulted in an increase of the fraction of erythromycin-resistant, edited cells to 99% (**Fig. 4c**). To determine if this increase is caused by the killing of non-edited cells, we compared the kanamycin-resistant cfu obtained after co-transformation of JEN53 cells with the CRISPR::ermAM(stop) or CRISPR::Ø constructs. We counted 5.3 times less kanamycin-resistant colonies after transformation of the ermAM(stop) construct ($2.5 \times 10^4/4.7 \times 10^3$; **Supplementary Fig. 8a**), a result that suggests targeting of a chromosomal locus by dual-RNA:Cas9 does, indeed, lead to the killing of non-edited cells. Finally, because the introduction of dsDNA breaks in the bacterial chromosome is known to trigger repair mechanisms that increase the rate of recombination of the damaged DNA[29], we investigated whether

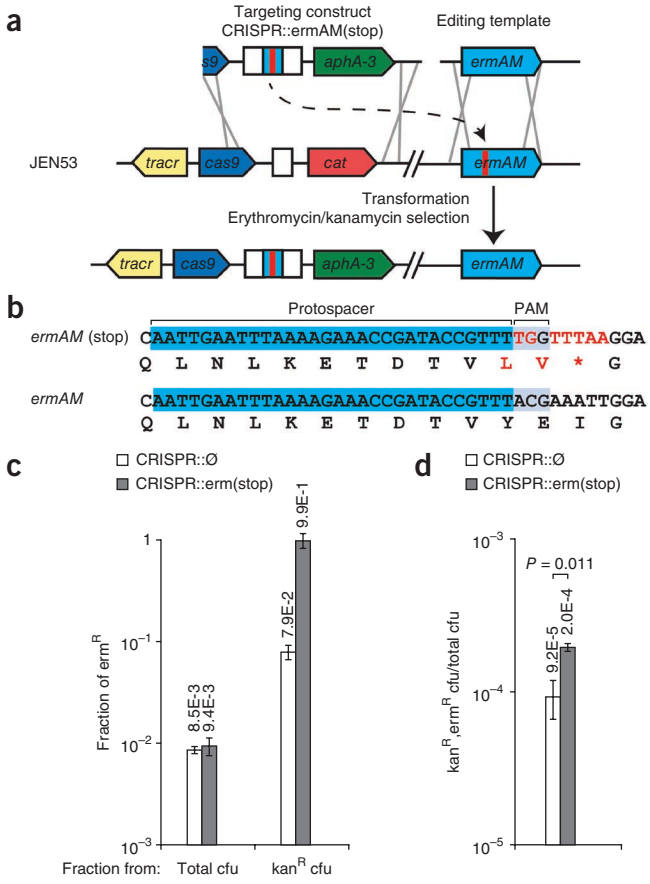

Figure 4 Mechanisms underlying editing using the CRISPR-Cas system. (**a**) A stop codon was introduced in the erythromycin resistance gene *ermAM* to generate strain JEN53. The wild-type sequence can be restored by targeting the stop codon with the CRISPR::ermAM(stop) construct, and using the ermAM wild-type sequence as an editing template. (**b**) Mutant and wild-type *ermAM* sequences. (**c**) Fraction of erythromycin-resistant (*erm*^{R}) cfu calculated from total or kanamycin-resistant (*kan*^{R}) cfu. (**d**) Fraction of total cells that acquire both the CRISPR construct and the editing template. Co-transformation of the CRISPR targeting construct produced more transformants (*t*-test, $P = 0.011$). Error bars, mean ± s.d. for three independent experiments.

cleavage by dual-RNA:Cas9 induces recombination of the editing template. We counted 2.2 times more colonies after co-transformation with the CRISPR::erm(stop) construct than with the CRISPR::Ø construct (**Fig. 4d**), indicating that there is a modest induction of recombination. Taken together, these results show that co-selection of transformable cells, induction of recombination by dual-RNA:Cas9-mediated cleavage and selection against non-edited cells contribute to the high efficiency of genome editing in *S. pneumoniae*.

As cleavage of the genome should kill non-edited cells, one would not expect to recover any cells that received the kanamycin resistance–containing targeting construct but not the editing template. However, even in experiments where the editing template was not introduced, we recovered many kanamycin-resistant colonies after transformation of the CRISPR::ermAM(stop) construct (**Supplementary Fig. 8a**). These cells that 'escape' CRISPR-induced death (CRISPR selection) produce a background that determines a limit of the method. This background frequency can be calculated as the ratio of CRISPR::ermAM(stop)/CRISPR::Ø cfu, 2.6×10^{-3} ($7.1 \times 10^1/2.7 \times 10^4$) in this experiment, meaning that if the recombination frequency of the editing template is less than this value, CRISPR selection will not efficiently recover the desired mutants above the background. To understand the origin of these cells, we genotyped eight background colonies and found that seven contained deletions of the targeting spacer (**Supplementary Fig. 8b**) and one harbored a presumably inactivating mutation in Cas9 (**Supplementary Fig. 8c**).

Genome editing with dual-RNA:Cas9 in *E. coli*

The activation of Cas9 targeting through the chromosomal integration of a CRISPR-Cas system is only possible in organisms that are highly recombinogenic. To develop a more general method that is applicable to other microbes, we decided to perform genome editing in *E. coli* using a plasmid-based CRISPR-Cas system. Two plasmids were constructed: a pCas9 plasmid carrying the tracrRNA, Cas9 and a chloramphenicol resistance cassette (**Supplementary Fig. 9**), and a pCRISPR kanamycin-resistant plasmid carrying the array of CRISPR spacers. To measure the efficiency of editing independently of CRISPR selection, we sought to introduce an A to C transversion in *rpsL* that confers streptomycin resistance[30]. We constructed a pCRISPR::rpsL plasmid harboring a spacer that would guide dual-RNA:Cas9 cleavage of the wild-type, but not the mutant *rpsL* allele (**Fig. 5a**). The pCas9 plasmid was first introduced into *E. coli* MG1655 and the resulting strain was co-transformed with the pCRISPR::rpsL plasmid and W542, an editing oligonucleotide containing the A to C mutation. We were only able to recover streptomycin-resistant colonies after transformation with the

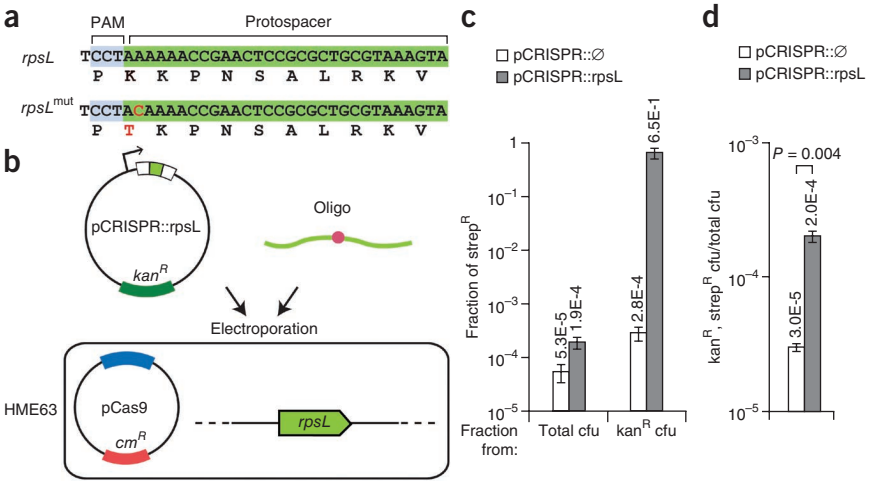

Figure 5 Genome editing with the CRISPR-Cas system in *E. coli*. (**a**) A K42T mutation conferring streptomycin resistance was introduced in *rpsL*. (**b**) A kanamycin-resistant plasmid carrying the CRISPR array (pCRISPR) targeting the gene to edit can be transformed in the HME63 recombineering strain containing a chloramphenicol-resistant plasmid harboring *cas9* and tracr (pCas9), together with an oligonucleotide specifying the mutation. (**c**) Fraction of streptomycin-resistant (*strep*^{R}) cfu calculated from total or kanamycin-resistant (*kan*^{R}) cfu. (**d**) Fraction of total cells that acquire both the pCRISPR plasmid and the editing oligonucleotide. Co-transformation of the pCRISPR targeting plasmid produced more transformants (*t*-test, $P = 0.004$). Error bars, mean ± s.d. for three independent experiments.

pCRISPR::rpsL plasmid, suggesting that cleavage by dual-RNA:Cas9 induces recombination of the oligonucleotide (**Supplementary Fig. 10**). However, the number of streptomycin-resistant colonies was two orders of magnitude lower than the number of kanamycin-resistant colonies, which are presumably cells that escape cleavage by dual-RNA:Cas9. Therefore, in these conditions, cleavage by dual-RNA:Cas9 facilitates the introduction of the mutation, but with an efficiency that is not enough to select the mutant cells above the background of 'escapers'.

To improve the efficiency of genome editing in *E. coli*, we applied our CRISPR-Cas system with recombineering, using dual-RNA:Cas9-induced cell death to select for the desired mutations. The pCas9 plasmid was introduced into the recombineering strain HME63 (ref. 31), which contains the Gam, Exo and Beta functions of the λ-red phage. The resulting strain was co-transformed with the pCRISPR::rpsL plasmid (or a pCRISPR::Ø control) and the W542 oligonucleotide (**Fig. 5b**). The recombineering efficiency was 5.3×10^{-5}, calculated as the fraction of total cells that became resistant to streptomycin when the control plasmid was used (**Fig. 5c**). In contrast, transformation with the pCRISPR::rpsL plasmid increased the percentage of mutant cells to 65 ± 14% (**Fig. 5c** and **Supplementary Fig. 2f**). We observed that the number of cfu was reduced by about three orders of magnitude after transformation of the pCRISPR::rpsL plasmid than the control plasmid ($4.8 \times 10^5/5.3 \times 10^2$; **Supplementary Fig. 11a**), suggesting that selection results from CRISPR-induced death of non-edited cells. To measure the rate at which cleavage by dual-RNA:Cas9 is inactivated, an important parameter of our method, we transformed cells with either pCRISPR::rpsL or the control plasmid without the W542 editing oligonucleotide (**Supplementary Fig. 11a**). This background of CRISPR 'escapers', measured as the ratio of pCRISPR::rpsL/pCRISPR::Ø cfu, was 2.5×10^{-4} ($1.2 \times 10^2/4.8 \times 10^5$). Genotyping eight of these escapers revealed that in all cases there was a deletion of the targeting spacer (**Supplementary Fig. 11b**). This background is higher than the recombineering efficiency of the *rpsL* mutation, 5.3×10^{-5}, which suggested that to obtain 65% edited cells, cleavage by dual-RNA:Cas9 must induce oligonucleotide recombination. To confirm this, we compared the number of kanamycin- and streptomycin-resistant cfu after transformation of pCRISPR::rpsL or pCRISPR::Ø (**Fig. 5d**). As in the case for *S. pneumoniae*, we observed a modest induction of recombination, about 6.7 fold ($2.0 \times 10^{-4}/3.0 \times 10^{-5}$). Taken together, these results indicate that the CRISPR-Cas system provides a method for selecting mutations introduced by recombineering.

DISCUSSION

Here we show that CRISPR-Cas systems can be used for targeted genome editing in bacteria by the co-introduction of a targeting construct that kills wild-type cells and an editing template that both eliminates cleavage by dual-RNA:Cas9 and introduces the desired mutations. Different types of mutations (insertions, deletions or scar-less single-nucleotide substitutions) can be generated. Multiple mutations can be introduced at the same time. The specificity and versatility of editing using the CRISPR-Cas system rely on several unique properties of the Cas9 endonuclease: (i) its target specificity can be programmed with a small RNA, without the need for enzyme engineering, (ii) target specificity is very high, determined by a 20-bp RNA-DNA interaction with low probability of nontarget recognition, (iii) almost any sequence can be targeted, the only requirement being the presence of an adjacent NGG sequence, (iv) almost any mutation in the NGG sequence, as well as mutations in the seed sequence of the protospacer, eliminates targeting.

We show that genome engineering using the CRISPR-Cas system works not only in highly recombinogenic bacteria such as *S. pneumoniae*, but also in *E. coli*. Results in *E. coli* suggest that the method should be applicable to other microorganisms for which plasmids can be introduced. In *E. coli*, the approach complements recombineering of mutagenic oligonucleotides. To use this methodology in microbes where recombineering is not possible, the host homologous recombination machinery could be used by providing the editing template on a plasmid. In addition, because accumulated evidence indicates that CRISPR-mediated cleavage of the chromosome leads to cell death in many bacteria[23,32,33] and archaea[34,35], it is possible to envision the use of endogenous CRISPR-Cas systems for editing purposes.

In both *S. pneumoniae* and *E. coli*, we observed that although editing is facilitated by a co-selection of transformable cells and a small induction of recombination at the target site by dual-RNA:Cas9 cleavage, the mechanism that contributes the most to editing is the selection against non-edited cells. Therefore the major limitation of the method is the presence of a background of cells that escape CRISPR-induced cell death and lack the desired mutation. We showed that these 'escapers' arise primarily through the deletion of the targeting spacer, presumably after the recombination of the repeat sequences that flank the targeting spacer. Future improvements will focus on the engineering of flanking sequences that can still support the biogenesis of functional crRNAs but that are sufficiently different from one another to eliminate recombination. Alternatively, the direct transformation of chimeric crRNAs[7] could be explored. In the particular case of *E. coli*, the construction of the CRISPR-Cas system is not possible if this organism is also used as a cloning host. We solved this issue by placing Cas9 and the tracrRNA on a different plasmid than the CRISPR array. The engineering of an inducible system could also circumvent this limitation.

Although new DNA synthesis technologies provide the ability to create any sequence cost effectively with a high throughput, it remains a challenge to integrate synthetic DNA in living cells to create functional genomes. Recently, the co-selection MAGE strategy was shown to improve the mutation efficiency of recombineering by selecting a subpopulation of cells that has an increased probability of achieving recombination at or around a given locus[36]. In this method, the introduction of selectable mutations is used to increase the chances of generating nearby nonselectable mutations. As opposed to the indirect selection provided by this strategy, the use of the CRISPR-Cas system makes it possible to directly select for the desired mutation and to recover it with a high efficiency. These technologies add to the toolbox of genetic engineers, who, using these tools and DNA synthesis, are likely to substantially advance our ability to both decipher gene function and manipulate organisms for biotechnological purposes. While our work was in review, two studies reported the use of CRISPR-Cas systems to engineer eukaryotic genomes[37,38], and two more accompany our paper[39,40]. It is expected that these crRNA-directed genome editing technologies will be broadly useful in the basic and medical sciences.

METHODS

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

Accession code. Fully sequenced pLZ12spec, GenBank: KC112384.

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

We thank V. Fischetti and C. Euler for plasmid pLZ12spec, D. Court for the HME63 strain, J. Kern for plasmid pKD46, the Rockefeller University Genomic Resource Center for technical support and J. Roberts for the MG1655 strain. D.B. is supported by a Harvey L. Karp Discovery Award and the Bettencourt Schuller Foundation. D.C. is supported by the Medical Scientist Training Program. F.Z. is supported by a US National Institutes of Health (NIH) Director's Pioneer Award (DP1MH100706),

Transformative R01, the Keck, McKnight, Gates, Damon Runyon, Searle Scholars, Klingenstein, and Simons Foundations, Bob Metcalfe, Mike Boylan and Jane Pauley. L.A.M. is supported by the Searle Scholars Program, the Rita Allen Scholars Program, a Irma T. Hirschl Award and a NIH Director's New Innovator Award (1DP2AI104556-01).

AUTHOR CONTRIBUTIONS

W.J., D.B. and L.A.M. designed the experiments; W.J., D.B. and D.C. performed experiments; W.J., D.B., F.Z. and L.A.M. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/doifinder/10.1038/nbt.2508.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.* **11**, 636–646 (2010).
2. Bogdanove, A.J. & Voytas, D.F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843–1846 (2011).
3. Stoddard, B.L. Homing endonuclease structure and function. *Q. Rev. Biophys.* **38**, 49–95 (2005).
4. Bae, T. & Schneewind, O. Allelic replacement in *Staphylococcus aureus* with inducible counter-selection. *Plasmid* **55**, 58–63 (2006).
5. Sung, C.K., Li, H., Claverys, J.P. & Morrison, D.A. An *rpsL* cassette, janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* **67**, 5190–5196 (2001).
6. Sharan, S.K., Thomason, L.C., Kuznetsov, S.G. & Court, D.L. Recombineering: a homologous recombination-based method of genetic engineering. *Nat. Protoc.* **4**, 206–223 (2009).
7. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
8. Deveau, H., Garneau, J.E. & Moineau, S. CRISPR-Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–493 (2010).
9. Horvath, P. & Barrangou, R. CRISPR-Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
10. Terns, M.P. & Terns, R.M. CRISPR-based adaptive immune systems. *Curr. Opin. Microbiol.* **14**, 321–327 (2011).
11. van der Oost, J., Jore, M.M., Westra, E.R., Lundgren, M. & Brouns, S.J. CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* **34**, 401–407 (2009).
12. Brouns, S.J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960–964 (2008).
13. Carte, J., Wang, R., Li, H., Terns, R.M. & Terns, M.P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **22**, 3489–3496 (2008).
14. Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).
15. Hatoum-Aslan, A., Maniv, I. & Marraffini, L.A. Mature clustered, regularly interspaced, short palindromic repeats RNA (crRNA) length is measured by a ruler mechanism anchored at the precursor processing site. *Proc. Natl. Acad. Sci. USA* **108**, 21218–21222 (2011).
16. Haurwitz, R.E., Jinek, M., Wiedenheft, B., Zhou, K. & Doudna, J.A. Sequence- and structure-specific RNA processing by a CRISPR endonuclease. *Science* **329**, 1355–1358 (2010).
17. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).
18. 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).
19. Makarova, K.S., Aravind, L., Wolf, Y.I. & Koonin, E.V. Unification of Cas protein families and a simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol. Direct* **6**, 38 (2011).
20. Barrangou, R. RNA-mediated programmable DNA cleavage. *Nat. Biotechnol.* **30**, 836–838 (2012).
21. Brouns, S.J. Molecular biology. A Swiss army knife of immunity. *Science* **337**, 808–809 (2012).
22. Carroll, D. A CRISPR approach to gene targeting. *Mol. Ther.* **20**, 1658–1660 (2012).
23. Bikard, D., Hatoum-Aslan, A., Mucida, D. & Marraffini, L.A. CRISPR interference can prevent natural transformation and virulence acquisition during in vivo bacterial infection. *Cell Host Microbe* **12**, 177–186 (2012).
24. Sapranauskas, R. *et al.* The *Streptococcus thermophilus* CRISPR-Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.* **39**, 9275–9282 (2011).
25. Semenova, E. *et al.* Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. USA* **108**, 10098–10103 (2011).
26. Wiedenheft, B. *et al.* RNA-guided complex from a bacterial immune system enhances target recognition through seed sequence interactions. *Proc. Natl. Acad. Sci. USA* **108**, 10092–10097 (2011).
27. Zahner, D. & Hakenbeck, R. The *Streptococcus pneumoniae* beta-galactosidase is a surface protein. *J. Bacteriol.* **182**, 5919–5921 (2000).
28. Marraffini, L.A., Dedent, A.C. & Schneewind, O. Sortases and the art of anchoring proteins to the envelopes of gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 192–221 (2006).
29. Motamedi, M.R., Szigety, S.K. & Rosenberg, S.M. Double-strand-break repair recombination in *Escherichia coli*: physical evidence for a DNA replication mechanism *in vivo*. *Genes Dev.* **13**, 2889–2903 (1999).
30. Hosaka, T. *et al.* The novel mutation K87E in ribosomal protein S12 enhances protein synthesis activity during the late growth phase in *Escherichia coli*. *Mol. Genet. Genomics* **271**, 317–324 (2004).
31. Costantino, N. & Court, D.L. Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc. Natl. Acad. Sci. USA* **100**, 15748–15753 (2003).
32. Edgar, R. & Qimron, U. The *Escherichia coli* CRISPR system protects from lambda lysogenization, lysogens, and prophage induction. *J. Bacteriol.* **192**, 6291–6294 (2010).
33. Marraffini, L.A. & Sontheimer, E.J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**, 568–571 (2010).
34. Fischer, S. *et al.* An archaeal immune system can detect multiple Protospacer Adjacent Motifs (PAMs) to target invader DNA. *J. Biol. Chem.* **287**, 33351–33363 (2012).
35. Gudbergsdottir, S. *et al.* Dynamic properties of the *Sulfolobus* CRISPR-Cas and CRISPR/Cmr systems when challenged with vector-borne viral and plasmid genes and protospacers. *Mol. Microbiol.* **79**, 35–49 (2011).
36. Wang, H.H. *et al.* Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* **9**, 591–593 (2012).
37. Cong, L. *et al.* Multiplex genome engineering using CRISPR-Cas systems. *Science* doi:10.1126/science.1231143 (3 January 2013).
38. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* doi:10.1126/science.1232033 (3 January 2013).
39. 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.* advance online publication, doi:10.1038/nbt.2507 (29 January 2013).
40. Hwang, W.Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* advance online publication, doi:10.1038/nbt.2501 (29 January 2013).

ONLINE METHODS

Strains and culture conditions. *S. pneumoniae* strain R6 (ref. 41) was provided by A. Tomasz. Strain crR6 was generated in a previous study[23]. Liquid cultures of *S. pneumoniae* were grown in THYE medium (30 g/l Todd-Hewitt agar, 5 g/l yeast extract). Cells were plated on tryptic soy agar (TSA) supplemented with 5% defibrinated sheep blood. When appropriate, antibiotics were added as followings: kanamycin (400 μg/ml), chloramphenicol (5 μg/ml), erythromycin (1 μg/ml), streptomycin (100 μg/ml) or spectinomycin (100 μg/ml). Measurements of β-galactosidase activity were made using the Miller assay as previously described[27].

E. coli strains MG1655 and HME63 (derived from MG1655, Δ(argF-lac) U169 λ cI857 Δcro-bioA galK tyr 145 UAG mutS<>amp)[31] were provided by J. Roberts and D. Court, respectively. Liquid cultures of *E. coli* were grown in LB medium (Difco). When appropriate, antibiotics were added as followings: chloramphenicol (25 μg/ml), kanamycin (25 μg/ml) and streptomycin (50 μg/ml).

***S. pneumoniae* transformation.** Competent cells were prepared as described previously[23]. For all genome editing transformations, cells were gently thawed on ice and resuspended in 10 volumes of M2 medium supplemented with 100 ng/ml of competence-stimulating peptide CSP1 (ref. 42), and followed by addition of editing constructs (editing constructs were added to cells at a final concentration of 0.7 ng/μl to 2.5 μg/μl). Cells were incubated 20 min at 37 °C before the addition of 2 μl of targeting constructs and then incubated 40 min at 37 °C. Serial dilutions of cells were plated on the appropriate medium to determine cfu.

***E. coli* Lambda-red recombineering.** Strain HME63 was used for all recombineering experiments. Recombineering cells were prepared and handled according to a previously published protocol[6]. Briefly, a 2 ml overnight culture (LB medium) inoculated from a single colony obtained from a plate was grown at 30 °C. The overnight culture was diluted 100-fold and grown at 30 °C with shaking (200 r.p.m.) until the A_{600} was 0.4–0.5 (~3 h). For Lambda-red induction, the culture was transferred to a 42 °C water bath to shake at 200 r.p.m. for 15 min. Immediately after induction, the culture was swirled in an ice-water slurry and chilled on ice for 5–10 min. Cells were then washed and aliquoted according to the protocol. For electro-transformation, 50 μl of cells were mixed with 1 mM of salt-free oligos (IDT) or 100–150 ng of plasmid DNA (prepared by QIAprep Spin Miniprep Kit, Qiagen). Cells were electroporated using 1 mm Gene Pulser cuvette (Bio-rad) at 1.8 kV and were immediately resuspended in 1 ml of room temperature LB medium. Cells were recovered at 30 °C for 1–2 h before being plated on LB agar with appropriate antibiotic resistance and incubated at 32 °C overnight.

Preparation of *S. pneumoniae* genomic DNA. For transformation purposes, *S. pneumoniae* genomic DNA was extracted using the Wizard Genomic DNA Purification Kit, following instructions provided by the manufacturer (Promega). For genotyping purposes, 700 μl of overnight *S. pneumoniae* cultures were pelleted, resuspended in 60 μl of lysozyme solution (2 mg/ml) and incubated 30 min at 37 °C. The genomic DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen).

Strain construction. All primers used in this study are provided in **Supplementary Table 2**. To generate *S. pneumoniae* crR6M, an intermediate strain, LAM226, was made. In this strain the *aphA-3* gene (providing kanamycin resistance) adjacent to the CRISPR array of *S. pneumoniae* crR6 strain was replaced by a *cat* gene (providing chloramphenicol resistance). Briefly, crR6 genomic DNA was amplified using primers L448/L444 and L447/L481, respectively. The *cat* gene was amplified from plasmid pC194 (ref. 43) using primers L445/L446. Each PCR product was gel-purified and all three were fused by SOEing PCR[44] with primers L448/L481. The resulting PCR product was transformed into competent *S. pneumoniae* crR6 cells and chloramphenicol-resistant transformants were selected. To generate *S. pneumoniae* crR6M, *S. pneumoniae* crR6 genomic DNA was amplified by PCR using primers L409/L488 and L448/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L409/L481. The resulting PCR product was used to transform competent *S. pneumoniae* LAM226 cells and kanamycin-resistant transformants were selected.

To generate *S. pneumoniae* crR6Rc, *S. pneumoniae* crR6M genomic DNA was amplified by PCR using primers L430/W286, and *S. pneumoniae* LAM226 genomic DNA was amplified by PCR using primers W288/L481. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was used to transform competent *S. pneumoniae* crR6M cells and chloramphenicol-resistant transformants were selected.

To generate *S. pneumoniae* crR6Rk, *S. pneumoniae* crR6M genomic DNA was amplified by PCR using primers L430/W286 and W287/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was used to transform competent *S. pneumoniae* crR6Rc cells and kanamycin-resistant transformants were selected.

To generate JEN37, *S. pneumoniae* crR6Rk genomic DNA was amplified by PCR using primers L430/W356 and W357/L481, respectively. Each PCR product was gel-purified and they were fused by SOEing PCR with primers L430/L481. The resulting PCR product was used to transform competent *S. pneumoniae* crR6Rc cells and kanamycin-resistant transformants were selected.

To generate JEN38, R6 genomic DNA was amplified using primers L422/L461 and L459/L426, respectively. The *ermAM* gene (specifying erythromycin resistance) was amplified from plasmid pFW15 (ref. 45) using primers L457/L458. Each PCR product was gel-purified and all three were fused by SOEing PCR with primers L422/L426. The resulting PCR product was used to transform competent *S. pneumoniae* crR6Rc cells and erythromycin-resistant transformants were selected.

S. pneumoniae JEN53 was generated in two steps. First JEN43 was constructed as illustrated in **Supplementary Figure 5**. JEN53 was generated by transforming competent JEN43 cells with genomic DNA of JEN25 and selecting on both chloramphenicol and erythromycin.

To generate *S. pneumoniae* JEN62, *S. pneumoniae* crR6Rk genomic DNA was amplified by PCR using primers W256/W365 and W366/L403, respectively. Each PCR product was purified and ligated by Gibson assembly. The assembly product was used to transform competent *S. pneumoniae* crR6Rc cells and kanamycin-resistant transformants were selected.

Plasmid construction. pDB97 was constructed through phosphorylation and annealing of oligonucleotides B296/B297, followed by ligation in pLZ12spec[46] digested by EcoRI/BamHI.

pDB98 was obtained after cloning the CRISPR leader sequence and was cloned together with a repeat-spacer-repeat unit into pLZ12spec. This was achieved through amplification of crR6Rc DNA with primers B298/B320 and B299/B321, followed by SOEing PCR of both products and cloning in pLZ12spec with restriction sites BamHI/EcoRI. In this way the spacer sequence in pDB98 was engineered to contain two BsaI restriction sites in opposite directions that allow for the scar-less cloning of new spacers.

pDB99 to pDB108 were constructed by annealing of oligonucleotides B300/B301 (pDB99), B302/B303 (pDB100), B304/B305 (pDB101), B306/B307 (pDB102), B308/B309 (pDB103), B310/B311 (pDB104), B312/B313 (pDB105), B314/B315 (pDB106), B315/B317 (pDB107), B318/B319 (pDB108), followed by ligation in pDB98 cut by BsaI.

The pCas9 plasmid was constructed as follow. Essential CRISPR elements were amplified from *S. pyogenes* SF370 genomic DNA with flanking homology arms for Gibson Assembly. The tracrRNA and Cas9 were amplified with oligos HC008 and HC010. The leader and CRISPR sequences were amplified HC011/HC014 and HC015/HC009, so that two BsaI type IIS sites were introduced in between two direct repeats to facilitate easy insertion of spacers.

pCRISPR was constructed by subcloning the pCas9 CRISPR array in pZE21-MCS1 through amplification with oligos B298+B299 and restriction with EcoRI and BamHI. The rpsL targeting spacer was cloned by annealing of oligos B352+B353 and cloning in the BsaI cut pCRISPR giving pCRISPR::rpsL.

Generation of targeting and editing constructs. Targeting constructs used for genome editing were made by Gibson assembly[47] of Left PCRs and Right PCRs (**Supplementary Table 3**). Editing constructs were made by SOEing PCR[42] fusing PCR products A (PCR A), PCR products B (PCR B) and

doi:10.1038/nbt.2508

PCR products C (PCR C) when applicable (**Supplementary Table 3**). The CRISPR::Ø and CRISPR::ermAM(stop) targeting constructs were generated by PCR amplification of JEN62 and crR6 genomic DNA respectively, with oligos L409 and L481.

Generation of targets with randomized PAM or protospacer sequences. The five nucleotides following the spacer 1 target were randomized through amplification of $R6^{8232.5}$ genomic DNA with primers W377/L426. This PCR product was then assembled with the *cat* gene and the *srtA* upstream region that were amplified from the same template with primers L422/W376. Eighty ng of the assembled DNA was used to transform strains R6 and crR6. Samples for the randomized targets were prepared using the following primers: B280-B290/L426 to randomize bases 1–10 of the target and B269-B278/L426 to randomize bases 10–20. Primers L422/B268 and L422/B279 were used to amplify the *cat* gene and *srtA* upstream region to be assembled with the first and last ten PCR products, respectively. The assembled constructs were pooled and 30 ng was used to transform R6 and crR6. After transformation, cells were plated on chloramphenicol selection. For each sample more than 2×10^5 cells were pooled together in 1 ml of THYE and genomic DNA was extracted with the Promega Wizard kit. Primers B250/B251 were used to amplify the target region. PCR products were tagged and run on one Illumina MiSeq paired-end lane using 300 cycles.

41. Hoskins, J. *et al.* Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J. Bacteriol.* **183**, 5709–5717 (2001).
42. Havarstein, L.S., Coomaraswamy, G. & Morrison, D.A. An unmodified heptadecapeptide pheromone induces competence for genetic transformation in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA* **92**, 11140–11144 (1995).
43. Horinouchi, S. & Weisblum, B. Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* **150**, 815–825 (1982).
44. Horton, R.M. In vitro recombination and mutagenesis of DNA: SOEing Together tailor-made genes. *Methods Mol. Biol.* **15**, 251–261 (1993).
45. Podbielski, A., Spellerberg, B., Woischnik, M., Pohl, B. & Lutticken, R. Novel series of plasmid vectors for gene inactivation and expression analysis in group A streptococci (GAS). *Gene* **177**, 137–147 (1996).
46. Husmann, L.K., Scott, J.R., Lindahl, G. & Stenberg, L. Expression of the Arp protein, a member of the M protein family, is not sufficient to inhibit phagocytosis of *Streptococcus pyogenes*. *Infect. Immun.* **63**, 345–348 (1995).
47. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).