Make-or-break prime editing for bacterial genome engineering

Monica Rengifo-Gonzalez[1,#], Maria-Vittoria Mazzuoli[1,#], Axel B. Janssen[1,#], Anne-Stéphanie Rueff[1], Jessica Burnier[1], Xue Liu[2,*], Jan-Willem Veening[1,*]

[1]Department of Fundamental Microbiology, Faculty of Biology and Medicine, University of Lausanne, Biophore Building, CH-1015 Lausanne, Switzerland

[2] Department of Pathogen Biology, Base for International Science and Technology Cooperation: Carson Cancer Stem Cell Vaccines R&D Center，International Cancer Center，Shenzhen University Medical School, Shenzhen 518055, Guangdong, China.

[#]These authors contributed equally

*Correspondence to Xue Liu (xueliu@szu.edu.cn) or Jan-Willem Veening (Jan-Willem.Veening@unil.ch), tel: +41 (0)21 6925625

Abstract

CRISPR-Cas9 has revolutionized genome engineering by allowing precise introductions of DNA double-strand breaks (DSBs). However, genome engineering in bacteria is still a complex, multi-step process requiring a donor DNA template for repair of DSBs. Prime editing circumvents this need as the repair template is indirectly provided within the prime editing guide RNA (pegRNA). Here, we developed make-or-break Prime Editing (mbPE) that allows for precise and effective genetic engineering in the opportunistic human pathogen *Streptococcus pneumoniae*. In contrast to traditional prime editing in which a nicking Cas9 is employed, mbPE harnesses wild type Cas9 in combination with a pegRNA that destroys the seed region or protospacer adjacent motif. Since most bacteria poorly perform template-independent end joining, correctly genome-edited clones are selectively enriched during mbPE. We show that mbPE is RecA-independent and can be used to introduce point mutations, deletions and targeted insertions, including protein tags such as a split luciferase, at selection efficiencies of over 93%. mbPE enables sequential genome editing, is scalable, and can be used to generate pools of mutants in a high-throughput manner. The mbPE system and pegRNA design guidelines described here will ameliorate future bacterial genome editing endeavors.

Introduction

The use of CRISPR/Cas9 technology has facilitated genetic engineering for a wide range of bacteria, including difficult or laborious to transform species such as *Staphylococcus aureus*, *Clostridium* spp. and *Streptomyces* spp.[1–5]. The basis for CRISPR/Cas9-based bacterial genome editing is the selective enrichment of transformants by targeting *Streptococcus pyogenes* Cas9 to the genetic location of interest by a single guide RNA (sgRNA) followed by production of a double-stranded break (DSB)[6]. Thus, cells in which the donor template DNA has not been incorporated through homologous recombination, will be targeted thereby enriching for edited clones[1,7,8]. This is particularly efficient in bacteria, as most strains lack (or poorly perform) template-independent end joining mechanisms such as non-homologous end joining (NHEJ) or alternative end joining[2,9,10]. Thus, only clones edited at the DSB and thereby disrupting the protospacer adjacent motif (PAM) or sgRNA seed region will survive. The major drawback of CRISPR/Cas9-based genome editing in bacteria is that several steps are involved such as the cloning of a sgRNA and the production or cloning of a template DNA[1,11].

Recently, prime editors such as PE2 were described in which a Cas9 nickase is fused to an engineered reverse transcriptase[7,12]. PE2 has the advantage that it does not introduce DSBs, which, in eukaryotes typically leads to error-prone end-joining via NHEJ and has the risk of inducing

chromosomal rearrangements[7]. In addition, prime editing uses a pegRNA for both targeting of the PE2 complex as well as a template for DNA repair in which the desired mutation is encoded. The pegRNA contains a spacer sequence to direct targeting of the prime editor to the PAM site, a primer binding site (PBS) that facilitates the hybridization of the 3′-end of the nicked DNA strand to the pegRNA and includes a reverse transcriptase template (RTT) that carries the intended genetic modification[12]. Because of these advantages, prime editing has rapidly gained traction and has been successfully applied to engineer the genomes in various eukaryotes including plants[13–16]. However, prime editing has not yet been widely applied for bacterial genome editing, probably because of limitations in the selection efficiency during the introduction of PE2-mediated mutations or due to the requirements of specific genetic backgrounds[17,18]. For instance, by mutating three 3′→5′ DNA exonucleases, PE efficiencies in *Escherichia coli* were improved 100-fold[18].

Here, we describe make-or-break prime editing (mbPE) as an efficient genome editing method in the opportunistic human pathogen *Streptococcus pneumoniae*. mbPE is a nuclease-based prime editing system that uses wild type *S. pyogenes* Cas9 fused to a bacterial codon-optimized reverse transcriptase under inducible control. Similar nuclease-based PE strategies have been employed in eukaryotic systems[19–21], but not yet in bacteria. Using mbPE, we were able to make point mutations, deletions over 100 bps as well as insertions up to 52 bps in a targeted fashion at high selection efficiencies. Using a pooled high-throughput approach, we established mbPE design rules and show that an optimal pegRNA, to insert a single adenine base, consists of a PBS of 16 nt and a RTT of 15 nt. We provide pegRNA vectors with different antibiotic selection markers allowing for sequential pneumococcal genome edits. The mbPE system developed here will be of great value to the pneumococcal research community and may function as a starting point for make-or-break genome editing approaches in other bacteria.

Results

Prime editing in *S. pneumoniae*

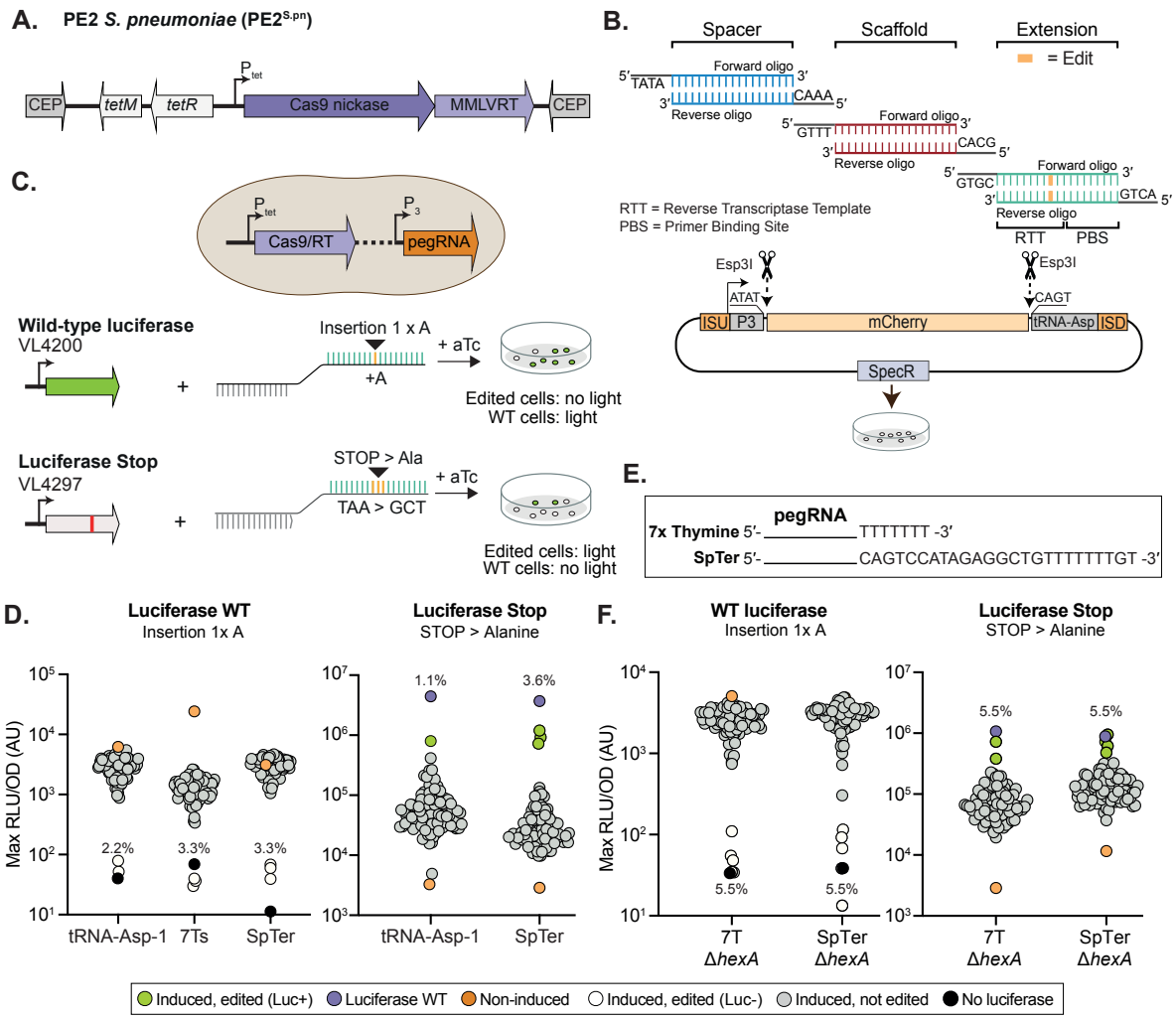
Several Cas9-based tools have been developed for *S. pneumoniae*[22–26], but prime editing has not been implemented yet. To establish a practical prime editing system for *S. pneumoniae*, we first focused on the prime editing tool PE2. PE2 is a fusion of the *S. pyogenes* Cas9^{H840A} nickase domain and an engineered Moloney murine leukemia virus reverse transcriptase domain (MMLVRT)[12]. Previously, we described tight inducible control by the addition of tetracycline-group molecules (including anhydrotetracycline (aTc)) of a catalytically inactive Cas9 (dCas9: D10A, H840A) expressed as a single copy from the non-essential chromosomal CEP locus (P_{tet}-*dcas9*) for CRISPR interference (CRISPRi)

83 in *S. pneumoniae*[25]. In a first step to a functioning PE2 for *S. pneumoniae*, we mutated the alanine
84 residue at position 10 of the P_{tet}-*dcas9* construct back to an aspartic acid to create P_{tet}-*cas9*^{H840A}. The
85 engineered MMLVRT domain and a linker domain from PE2 were codon optimized for optimal
86 expression in low GC-rich bacteria such as *S. pneumoniae* and cloned to the C-terminus of the Cas9
87 nickase resulting in strain VL4199 (P_{tet}-*cas9*^{H840A}-MMLVRT or P_{tet}-PE2$^{S.pn}$) (Fig. 1A).

88 Next, a pegRNA construction and *S. pneumoniae* integration vector was designed in which the
89 pegRNA is constitutively expressed from the constitutive P3 promoter including its +1 adenine
90 transcription start site to ensure that all cloned pegRNAs are efficiently transcribed and do not contain
91 undesirable uridines at the 5′ end[26–28]. For easy cloning, an *mCherry* cassette flanked by Esp3I
92 (isoschizomer of BsmBI) sites is present allowing for rapid, oligo-based, Golden Gate cloning of
93 pegRNAs (Fig. 1B). Sequence elements facilitating one-step library preparation for NGS were also
94 introduced[29] resulting in pegRNA cloning vectors pVL4132, pVL4133 and pVL4393 (carrying 7
95 thymine nucleotides, a *S. pneumoniae* terminator or a *S. pneumoniae tRNA-asp-1* after the extension
96 sequence, respectively). pegRNAs consist of four elements: 1) a 20 nt spacer sequence for targeting of
97 the prime editor, 2) a scaffold sequence for binding to Cas9, 3) a primer binding site (PBS) that allows
98 the 3′-end of the nicked DNA strand to hybridize to the pegRNA, and 4) a reverse transcriptase template
99 (RTT) containing the desired edit[12]. Since bacterial transcription termination leaves nucleotides at the
100 3′end of the pegRNA, which may interfere with efficient prime editing downstream, we cloned *S.*
101 *pneumoniae tRNA-asp-1* directly following the pegRNA (Fig. 1B). This strategy has been previously
102 used to produce multiple individual sgRNAs from a single polycistronic operon as the endogenous
103 tRNA-processing system precisely cleaves both ends of the *tRNA-asp-1* precursor[30]. To test whether
104 the PE2$^{S.pn}$ is functional, we designed several pegRNAs using PrimeDesign[31] to introduce either a
105 premature stop codon in the firefly luciferase gene (*luc*) expressed in pneumococcal strain VL4200, or
106 to introduce a repairing mutation in strain VL4297 in which *luc* has a premature stop codon (Fig. 1C).
107 Interestingly, induction of PE2$^{S.pn}$ together with a targeting pegRNA did not result in significant growth
108 delay, suggesting that *S. pneumoniae* can tolerate and readily repair ssDNA breaks (Fig. S1A). As
109 shown in Fig. 1D, approximately 2% of clones in which PE2$^{S.pn}$ was induced demonstrated the desired
110 bioluminescence phenotype. Sanger sequencing of several individual clones confirmed the *luc*
111 expression observations, demonstrating that bioluminescence is a good proxy for DNA editing (Fig.
112 S1B and S1C).

113 Recently, it was shown that a structured 3′-end of a pegRNA increases stability and improves
114 editing efficiency in human cells[32]. In addition, a functional mismatch repair system was shown to
115 negatively affect editing efficiencies[33,34]. To test whether the addition of a structured RNA motif at the
116 3′-end would also improve PE2 editing efficiency in *S. pneumoniae*, we replaced the tRNA sequence
117 for the structured and highly efficient *S. pneumoniae rpsT* transcriptional terminator[35] (SpTer) or a

118 stretch of 7 thymine nucleotides (7 Ts) (Fig. 1E). As shown in Fig. 1D, luminescence data of both
119 pegRNAs suggests improved efficiency of inserting a single A in *luc* (~3%) through PE2$^{S.pn}$-mediated
120 editing, compared to the tRNA-Asp1 system. Sanger sequencing of several bioluminescent-negative
121 clones confirmed the microtiter plate assay results (Fig. S1). Repairing an internal stop codon (TAA to
122 GCT) in *luc* was approximately 3 times more efficient with the SpTer terminator compared to the tRNA-
123 Asp1 construct (Fig. 1D). Deleting the pneumococcal mismatch repair gene *hexA*[36] further enhanced
124 PE2$^{S.pn}$ editing efficiency in the constructs with either the 7T or SpTer terminators (Fig. 1F). These
125 results demonstrate that prime editing using PE2$^{S.pn}$ in combination with a structured pegRNA works in
126 *S. pneumoniae* and can be used to incorporate insertions and mutations, albeit at a relatively low
127 selection efficiency.



128

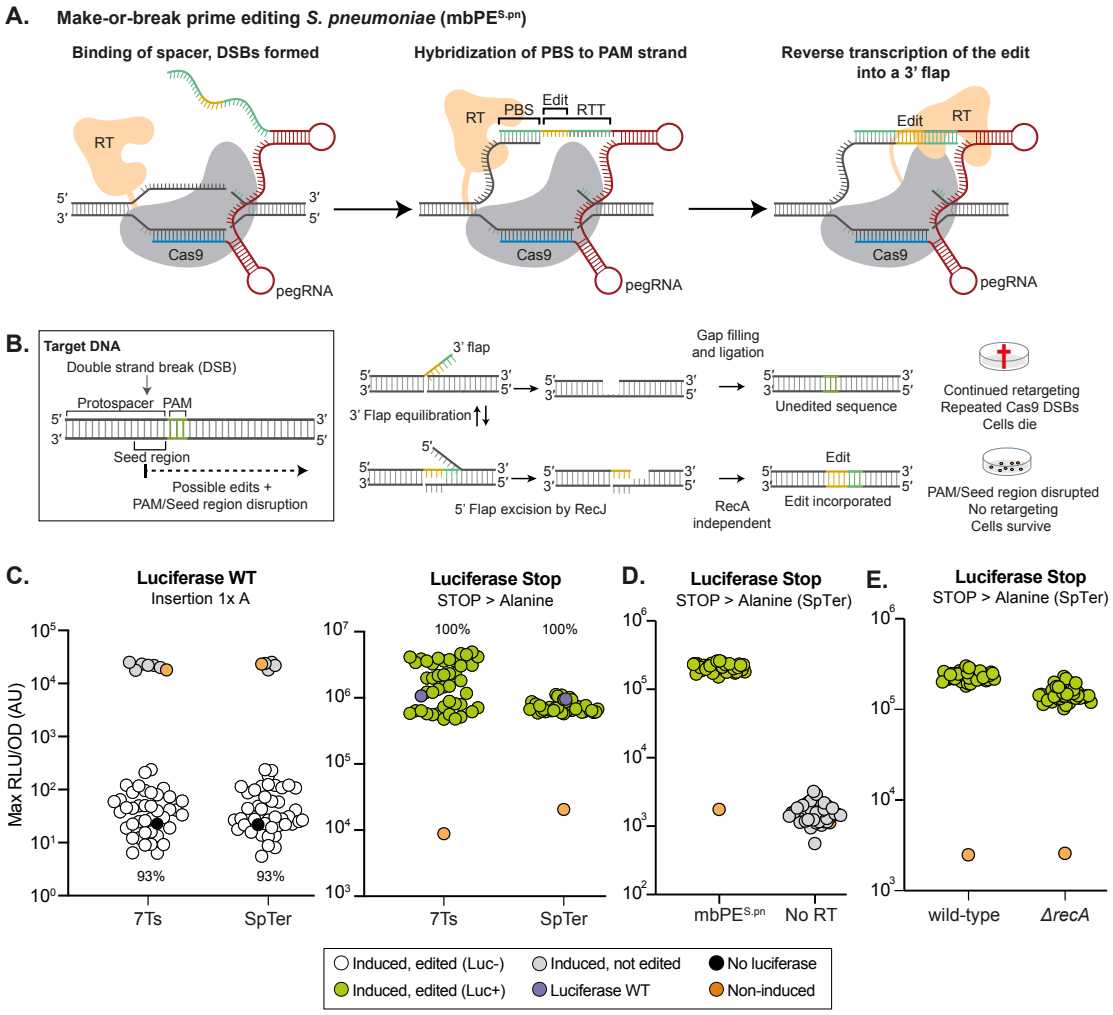
129 **Figure 1. Prime editing 2 (PE2) in *S. pneumoniae*.** (**A**) Schematic overview of PE2$^{S.pn}$. (**B**) Schematic overview
130 of the general cloning strategy of pegRNAs (shown is the *tRNA-asp-1* containing vector pVL4393 as backbone).
131 (**C**) Editing strategy used to test efficiency of PE2$^{S.pn}$ in pneumococcal strains expressing either a functional
132 luciferase (VL4200) or a *luc* gene interrupted by a premature stop codon (VL4297). Single colonies are picked

133 and grown in microtiter plates in the presence of luciferase substrate and bioluminescence is recorded. (**D**)
134 Induction of PE2$^{S.pn}$ by aTc in combination with expression of a pegRNA without 3′ leader, by processing of the
135 *tRNA-asp-1*, leads to accurate gene editing in approximately 2% of *S. pneumoniae* cells. Induction of PE2$^{S.pn}$ in
136 combination with 3′ extended pegRNAs provides an editing efficiency of approximately 3%. Each dot represents
137 a randomly picked clone grown in C+Y medium containing luciferin substrate. The maximal bioluminescence
138 (relative light units; RLU) reached, normalized by the optical density (OD), of each clone is shown (AU, arbitrary
139 units). RLU/OD values were normalized against the baseline level of medium luminescence obtained for each
140 experiment. A typical experiment is shown (repeated at least 3 times) with at least 90 colonies selected. (**E**)
141 Schematic overview of pegRNAs containing the *S. pneumoniae rpsT* terminator (SpTer) or a stretch of 7 Ts. (**F**)
142 Prime editing is slightly enhanced in the absence of mismatch repair (Δ*hexA*) in *S. pneumoniae*. Clones with either
143 an RLU/OD value below 2 x 10^2 (left), or an RLU/OD value above 7 x 10^6 (right), were considered to be edited
144 (as also validated by Sanger sequencing, Fig. S1).
145

146 **Make-or-break prime editing (mbPE) in *S. pneumoniae***

147 Since prime editing depends on successful hybridization of the 3′end of the nicked DNA to the
148 pegRNA[12], we wondered if prime editing would also be functional when, instead of a nick of the PAM-
149 containing strand, a DSB is introduced. Recent work in human HEK293T cells showed efficient prime
150 editing utilizing Cas9-induced DSBs that largely depend on NHEJ[19–21]. To test if DSB-driven prime
151 editing (here after called make-or-break prime editing: mbPE) would also work in bacteria lacking
152 NHEJ, we first mutated the alanine residue at position 840 within the Cas9 nickase moiety of PE2$^{S.pn}$
153 back to a histidine. This led to full restoration of wild type Cas9, resulting in strain VL5149 (P_{tet}-*cas9*-
154 MMLVRT or mbPE$^{S.pn}$) (Fig. 2A-B). This construct was combined with *luc* expression (strain VL5155)
155 and the same set of pegRNAs containing the SpTer or 7 Ts as previously used to verify the efficiency
156 of the nickase-based PE2$^{S.pn}$ (Fig. 1E). All pegRNAs were designed to either disrupt the PAM or the
157 seed region within *luc*. Thus, only correctly edited bacteria are expected to survive the expression of
158 mbPE$^{S.pn}$, as unedited repairs will lead to continued DSBs, hence the name make-or-break prime editing.
159 Induction of mbPE$^{S.pn}$ with aTc, in the presence of a targeting pegRNA, caused severe growth defects
160 indicating that DSBs were formed (Fig. S2A). The system is tightly inducible as reported previously[37],
161 as we observed a similar number of colonies compared to a strain that contained a non-targeting
162 pegRNA plated in the absence of aTc (Fig. S2B), suggesting that DSBs are only generated in the
163 presence of inducer. Strikingly, surviving colonies (approximately 1 - 5%, depending on the pegRNA
164 and aTc concentration), predominantly displayed the correct luciferase phenotype (93% of selected
165 colonies) (Fig. 2C). Sanger sequencing demonstrated that the precise edit was made as encoded by the
166 pegRNA template (Fig. S2C-F). Colonies that did not display the correct *luc*-edit either had suppressor
167 mutations in the Cas9 domain, or deletions in the P_{tet} promoter driving the expression of mbPE$^{S.pn}$ (Fig.
168 S2G). When comparing an sgRNA targeting *luc* with a pegRNA encoding the same spacer sequence,

169 we observed an increase in recovered colonies when the pegRNA was used, in line with the editing
170 efficiencies observed with PE2 (Fig. S2B). Note that the reverse transcriptase domain was essential for
171 prime editing since we did not observe any correctly edited clones when only the Cas9 domain was
172 expressed (Fig. 2D). PacBio HiFi long read whole genome sequencing of four edited clones did not
173 reveal signs of off-target editing nor any genomic rearrangements (Fig. S3B, Fig. S3C) (SRA accession
174 number PRJNA1062582).



175

176 **Figure 2. Make-or-break prime editing selectively enriches correctly edited clones.** (**A**) Schematic overview
177 of mbPE$^{S.pn}$. Binding of the spacer sequence to the target region leads to formation of DSBs by the Cas9 domain.
178 The PBS can hybridize to its complementary sequence. The reverse transcriptase domain of mbPE$^{S.pn}$
179 subsequently synthesizes cDNA using the RTT as template. (**B**) Schematic depicting the mechanism leading to
180 selective survival of edited clones. The hybridization of the 3′-flap containing the edit is in equilibrium with the
181 unedited strand. Repair with the edited region will prevent repeated DSBs by Cas9. (**C**) Induction of mbPE$^{S.pn}$ by
182 aTc in combination with expression of a structured pegRNA leads to accurate gene editing. Approximately 93%
183 of surviving *S. pneumoniae* cells after induction are accurately edited. Cells were grown in C+Y medium
184 containing luciferin substrate. The maximal bioluminescence reached of each clone is shown (RLU: relative light
185 units. OD: optical density). In these experiments, clones with either an RLU/OD value below 2 x 10^3 (left), or an
186 RLU/OD value above 2 x 10^5 (right), were considered to be edited (as also validated by Sanger sequencing). D)
187 In the absence of reverse transcriptase (strain VL7598, P_{tet}-*cas9*), no induced clones were edited, as indicated by

188 the lack of bioluminescence. E) mbPE$^{S.pn}$ in *recA* strain showing that RecA-dependent homologous
189 recombination is not required for editing.

190

191 **mbPE does not require RecA-dependent homologous recombination**

192 In Fig. 2D we established that reverse transcription of the pegRNA is required for correct genome
193 editing. DNA insertion in bacteria can occur by specific bacteriophage recombinases but typically is
194 established by RecA-dependent homologous recombination[23,38]. However, RecA typically requires long
195 homology regions. To test whether RecA is indeed not required for make-or-break prime editing in *S.*
196 *pneumoniae*, in which only short regions of homology are presented within the pegRNA, we replaced
197 *recA* with an antibiotic resistance cassette in strain VL7487 (mbPE$^{S.pn}$, pegRNA-luc+SpTer). Note that,
198 we were not able to construct this strain at 37°C but could efficiently construct it at 30°C. After inducing
199 the mbPE system and picking 42 surviving colonies on aTc plates, all demonstrated the encoded *luc*
200 repair in the *recA* strain (Fig. 2E). This data is in line with the original model for prime editing[12] in
201 which a 5′→3′ ssDNA exonuclease (RecJ in *S. pneumoniae*) would excise the 5′ flap followed by gap
202 filling by DNA polymerase and subsequent ligation by DNA ligase (Fig. 2B), a process that would not
203 require RecA-dependent homologous recombination.

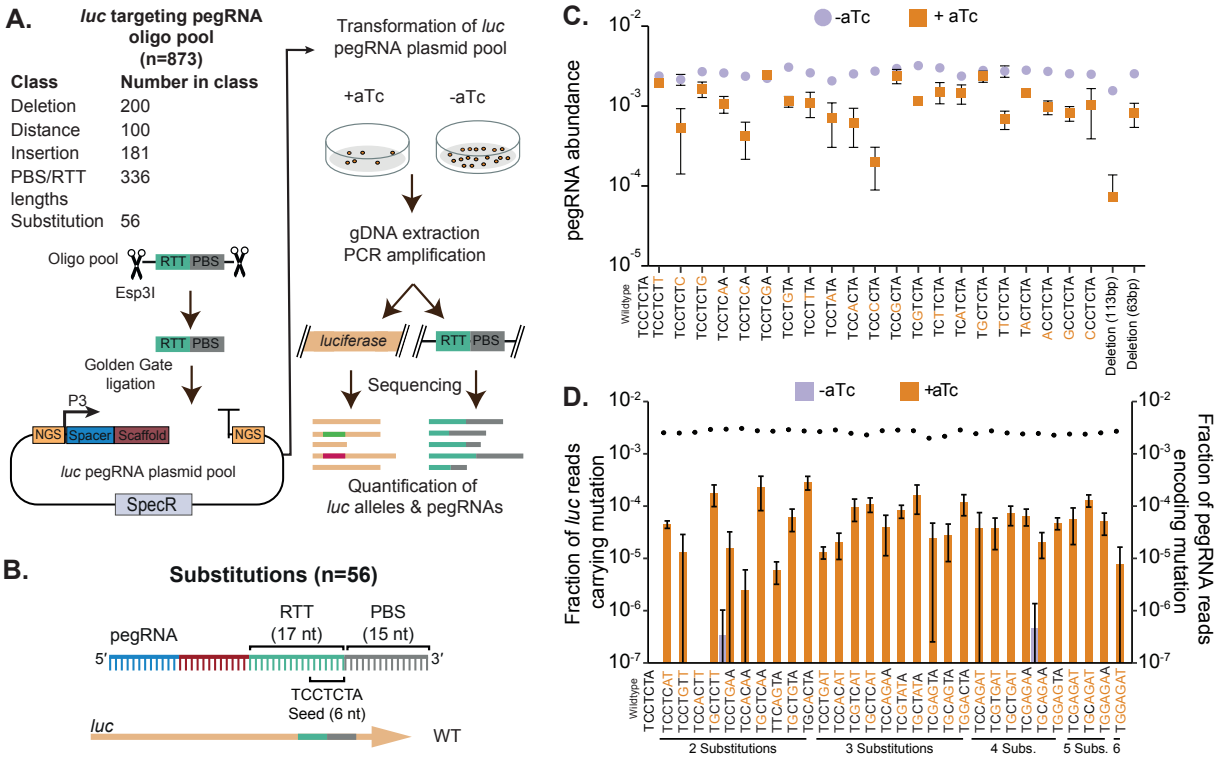
204

205 **Scalable pooled pegRNA library construction shows efficient installment of base pair**
206 **substitutions**

207 To explore the potential of mbPE$^{S.pn}$, we generated a pool of different pegRNAs targeting *luc* (Table
208 S1), followed by Golden Gate cloning (Fig. 3A). A total of 873 oligos were designed and synthesized
209 as a pool, containing different PBS and RTT sequences to introduce various genome edits. The pegRNA
210 pool was cloned into plasmid pVL4134 that already contains the *luc*-targeting spacer sequence. The
211 resulting pool was subsequently used to transform strain VL5155 carrying *luc* and the aTc-inducible
212 mbPE$^{S.pn}$ (see Methods). The library was plated on agar lacking or containing 20 ng/ml aTc to induce
213 mbPE$^{S.pn}$. Quadruplicate induced and uninduced experiments, each with a minimum of 40000 colonies,
214 were pooled, and their pegRNA contents and edited *luc* region were sequenced (Fig. 3A). In the
215 uninduced pool, all but 12 of the designed pegRNAs were present, and we did not observe some of the
216 pegRNAs encoding the longest insertions. Strikingly, most pegRNAs were also present in the induced
217 pool albeit at varying abundances. By counting and comparing the number of reads coming of each
218 pegRNA in both conditions, we could estimate a selection efficiency score for each pegRNA for a
219 specific mutation (Fig. 3A). By also sequencing the *luc* gene, we confirmed that most of the designed
220 edits were correctly made in the pool of mutants (see Methods). We note that while comparing the ratios

221 of reads of pegRNAs is useful to assess whether a certain edit is in principle possible, particularly when
222 the pegRNAs encode the same spacer and PBS, it does not quantitatively inform on editing efficiencies
223 as selection efficiencies will be very different depending on the level by which the pegRNA-encoded
224 mutation in the RTT affects targeting or editing. For instance, some pegRNAs might be less efficient at
225 cutting, potentially due to improper folding, which in turn might enhance editing efficiency since the
226 RT moiety has more time to produce the cDNA before a DSB is made.

227 First, we examined whether mbPE$^{S.pn}$ was able to install various substitutions (a>t, a>c, a>g, t>a, t>c,
228 t>g, c>a, c>t, c>g) of targets within a seed region. We designed pegRNAs that have the same 20 nt *luc*
229 spacer and 15 nt PBS, but varying RTTs encoding the different substitutions (and the necessary *luc* seed
230 disruptions) (Fig. 3B, Table S1). As shown in Fig. 3C, pegRNAs encoding all substitutions were present
231 at high frequencies under induced conditions. Next, we wondered if mbPE$^{S.pn}$ could introduce multiple
232 substitutions simultaneously. Indeed, all tested substitutions, up to the 6 adjacent substitutions tested
233 (CCTCTA> GGAGAT), were present in the induced library (Fig. 3D). For single base pair
234 substitutions, selection efficiency was more efficient when the edit was made to destroy the PAM and
235 less efficient when the edit was distal from the PAM (Fig. 4A).



236

237 **Figure 3. mbPE$^{S.pn}$ is highly scalable and efficient in making base pair substitutions. (A)** Cloning strategy to
238 produce a library of pegRNAs targeting *luc* from a ssDNA oligo pool. The oligo pool was amplified by PCR as
239 well as the backbone of pVL4134 that includes the *luc* spacer sequence and assembled by Golden Gate
240 digestion/ligation followed by transformation to *S. pneumoniae* containing *luc* and mbPE2$^{S.pn}$. Chromosomal
241 DNA was isolated from colonies grown in absence or presence of aTc (to induce mbPE2$^{S.pn}$) and PCR amplified

using staggered primers followed by illumina sequencing (see Methods). 2FAST2Q[39] was used to extract and count pegRNAs and *luc* alleles. (**B**) Schematic overview of the pegRNA design encoding for different base pair substitutions. (**C-D**) Single base pair substitutions and multiple base pair substitutions are installed with high selection efficiency. (**C**) pegRNA abundance of - aTc samples (control) and + aTc samples (orange squares) are shown. (**D**) The fraction of *luc* reads carrying the desired mutation over the total reads in that sample is shown for induced (aTc treated) and non-induced (control) samples on the left Y-axis. The average fraction of pegRNA reads for the substitutions is shown as dots on the right Y-axis.

Efficient DNA insertion and deletion with mbPE

After establishing that mbPE$^{S.pn}$ is highly effective at selecting clones containing point mutations, we examined the capacity of mbPE$^{S.pn}$ for DNA insertions or deletions (Fig. 4A, 4C). To test DNA insertion efficiencies, we designed 181 different insertion pegRNAs targeting *luc* with a constant 20 nt spacer, a constant 15 nt PBS, a constant 17 nt RTT scaffold followed by an RTT insertion of varying length (between 1 and 181 nt) that would insert in the *luc* seed region thereby preventing Cas9 retargeting (Fig. 4A, Table S1). We analyzed pegRNA abundances of uninduced and induced libraries, as well as sequenced the *luc* genes in the pools. We observed insertions up to 52 bps, indicating that there is a maximum length that can be installed through mbPE in *S. pneumoniae* (Fig. 4B). Note that longer pegRNAs were less well represented in the insertion pegRNA pool likely because of reduced abundance due to uneven synthesis and/or cloning efficiencies (Fig. 4B). This could have influenced our ability to observe the mutations encoded in these larger pegRNAs.

A similar strategy was employed to assess deletion efficiency inside *luc*, and 200 deletion pegRNAs were designed each containing a constant 20 nt *luc* spacer and constant 15 nt PBS. The RTT′s were 15 nt long, each shifted by one base pair relative to each other to create varying deletion sizes in *luc* whilst simultaneously disrupting the seed region (Fig. 4C). We observed a decrease in editing efficiency as deletion length increased, with the longest deletion observed being 184 bp (Fig. 4D). Indeed, we could readily obtain deletions at high selection efficiency when reconstructing a set of deletion pegRNAs encoding the 10, 63 or 127 base pair deletion (ranging between 80-90% efficiency). However, only 3 out of 28 tested clones showed reduced luciferase activity with the 184 bp deletion pegRNA and Sanger sequencing indicated that these were not correctly edited at the 184 bp position (Fig. 4E).

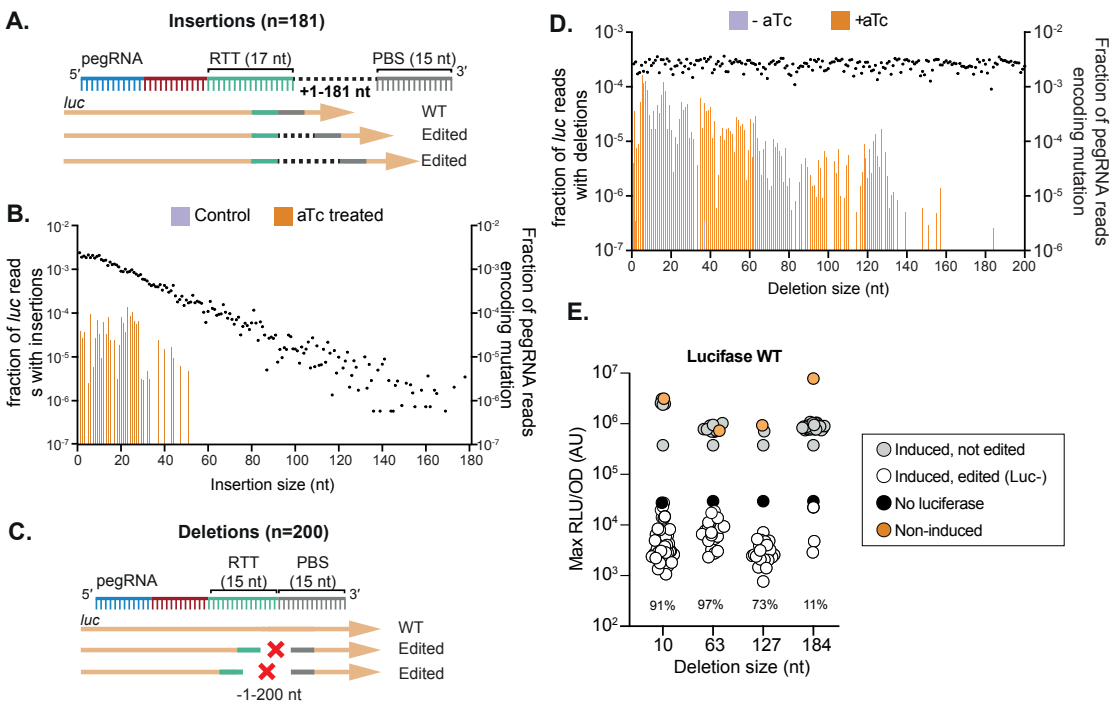
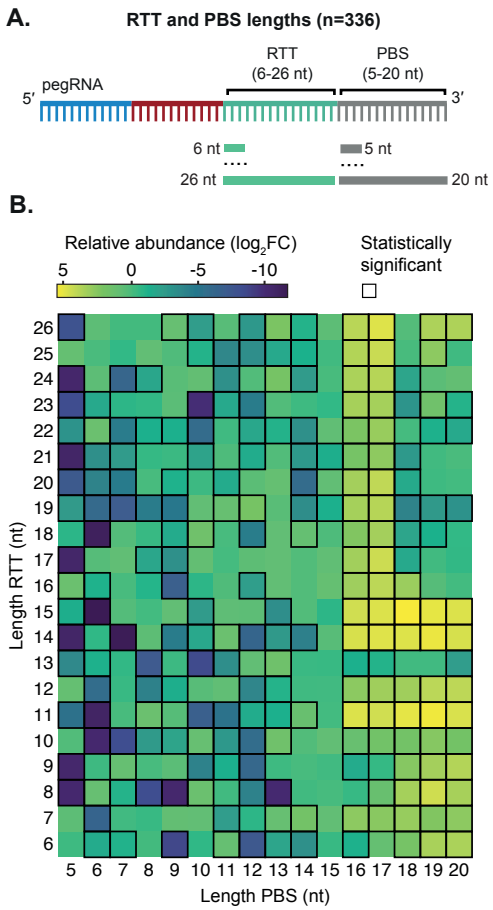


Figure 4. mbPE for DNA insertions and genome deletions. (**A**, **C**) Schematic overview of the approach used to test insertion and deletion efficiency in *luc* using mbPE. For pegRNAs encoding insertions, additional nucleotides were added between the RTT and PBS. For deletions, the RTT portion of the pegRNA was progressively moved within the *luc* gene, increasing deletion size, whilst the RTT remained constant. (**B**, **D**) Possibility of making insertions (**B**), and deletions (**D**) using mbPE. For each modification size, the fraction of *luc* reads carrying that mutation over the total reads in that sample, is displayed in bars on the left axis with bars, whilst the fraction of pegRNA reads from uninduced samples carrying the mutation is displayed with dots mapping to the right axis. (**E**) Verification of the ability to make deletions in *luc*, through targeted deletion of 10, 63, 127, or 184 bp. Clones with an RLU/OD value below 10^5 were considered to be edited (as also validated by Sanger sequencing).

pegRNA design rules for optimal mbPE

Previous work in human cells has shown that PE2 gene editing efficiencies highly depend on RTT and PBS length[12]. To determine the optimal RTT and PBS length for mbPE$^{S.pn}$ to introduce a single adenine nucleotide insertion in *S. pneumoniae*, we designed 336 pegRNAs, each encoding the same 20-nucleotide *luc* spacer sequence but varying in RTT length (6-26 nucleotides) and in PBS length (5-20 nucleotides). Each RTT contained the insertion of a single adenine nucleotide to disrupt the *luc* PAM, thereby preventing retargeting of Cas9 (Fig. 5A). Differential enrichment analysis of pegRNA abundance with and without mbPE$^{S.pn}$ inducer showed relatively more surviving clones with a PBS longer than 16 nts, and an RTT longer than 11. The most enriched pegRNAs for inserting a single adenine in *luc* contained a combination of a RTT length of 14 or 15 with a PBS length of 16 or 17 nts

294 (Fig. 5B). Together, these experiments provide guidelines for optimal pegRNA design and shows that
295 mbPE is a versatile editing tool capable of introducing single point mutations, multiple simultaneous
296 substitutions, insertions and deletions.



297

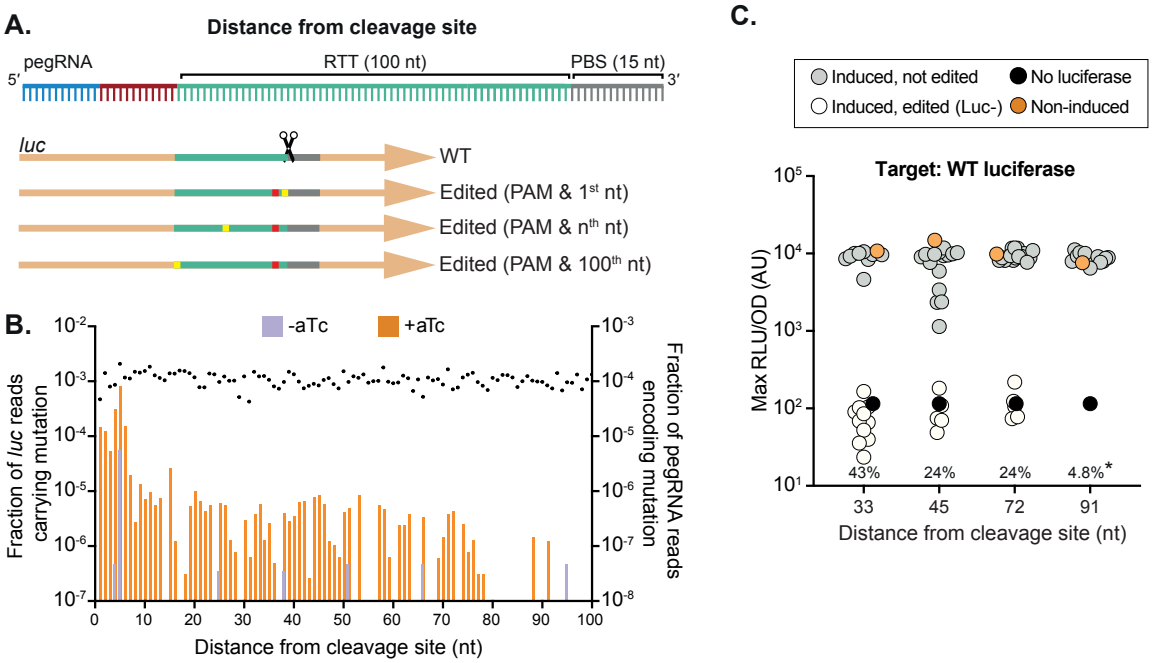
298 **Figure 5. pegRNA design rules for mbPE$^{S.pn}$.** (**A**) Schematic overview of the approach used to test mbPE$^{S.pn}$
299 efficiency at inserting a single adenine in *luc*. (**B**) Combinations of RTT and PBS with statistically significant
300 differences in pegRNA abundance between uninduced vs induced are outlined and overrepresented combinations
301 show higher log2 fold changes ($\log_2$FC) and are colored bright yellow. Most overrepresented pegRNAs of the
302 induced library compared to uninduced suggest that a 16 or 17 nts for the PBS and 14 or 15 nts for the RTT yield
303 the highest editing efficiency. Statistical significance was defined as an $|\log_2 FC| > 1$, and an adjusted p-value <
304 0.05.
305

306 **Exploring the editing distance from the seed region**

307 The strategies employed in this study guide the Cas9 nuclease moiety of mbPE$^{S.pn}$ to cut within the seed
308 region near the PAM motif or in the PAM as dictated by the spacer sequence, while the reverse
309 transcriptase (RT) component of mbPE$^{S.pn}$ introduces the desired genome edit. This approach prevents
310 Cas9 from retargeting the same location. The dependency on a PAM site is, however, a possible
311 limitation of the mbPE system. For mbPE to be broadly applicable for the introduction of mutations,
312 even in genomic regions without a nearby PAM sequence, editing should also be possible at sites distal

313 from the PAM. To determine the maximum distance from the Cas9 cut site at which a single nucleotide
314 substitution is possible, we designed a set of 100 pegRNAs with identical spacers. Each pegRNA
315 contained RTT sequences to mutate the PAM (preventing retargeting) while introducing a SNP at
316 varying distances from the PAM site (1-100 nts from the restriction site; Fig. 6A). As shown in Fig. 6B,
317 edits were observed as far as 91 bp from the PAM site.

318 Several pegRNAs were individually cloned, each with a mutation that destroys the PAM to
319 insert a silent mutation in *luc* and a substitution further from the PAM site. This showed that SNPs
320 could be efficiently introduced at a distance of 33 base pairs from the PAM (Fig. 6C). Editing was also
321 possible at further distances, but with increasingly reduced efficiency, similar to what was observed in
322 the pooled screen (Fig. 6B).



323

324 **Figure 6. mbPE can introduce edits at sites distal from the PAM.** (**A**) Schematic overview of the RTT design
325 strategy used to edit bases distal from the PAM site. Every pegRNA introduces a single PAM mutation (G to C)
326 and a second, purine to pyrimidine (or vice versa) transversion introduced at varying distances from the PAM
327 sequence. (**B**) Fraction of correctly edited *luc* alleles and pegRNA read counts over the total reads in that sample
328 of DNA extracted from cells grown in the absence (control) or presence of aTc. The average fraction of pegRNA
329 reads for the substitutions is shown as dots on the right Y-axis. Edits are observed up to 91 bp from the PAM site.
330 (**C**) Validation of the maximum editing distance through individually cloned pegRNAs. The mutations conferred
331 by the pegRNAs all encode the G to C PAM mutation and a second mutation at varying distances. The 33, and 45
332 bps substitutions from the PAM site, confer a premature stop codon, whilst the pegRNAs encoding mutation either
333 72 or 91 bps from the PAM site resulted in a glycine to cysteine (72 bps) or isoleucine to asparagine (91 bps)
334 substitution. The percentages indicate the fraction of clones that were phenotypically observed to be edited. Clones

335 with an RLU/OD value below 3 x 10^3 were considered to be edited. The mutation at 91 nts (*) could only be
336 confirmed by sequencing individual clones, as the amino acid change does not affect luciferase function. This
337 showed an editing efficiency of 4.8 % (1 out of 21 tested colonies) with the correct genome edit.
338

339 **mbPE to introduce a split-*luc* tag for protein-protein interactions**

340 Our pooled pegRNA assays indicated that DNA up to 52 bps can be inserted into the pneumococcal
341 genome using mbPE (Fig. 4B). To test whether mbPE could be used to introduce protein tags, we
342 considered the split luciferase tag. The use of split luciferase has emerged as a powerful approach to
343 identify protein-protein interactions in live bacterial cells in which the bait is fused to the large luciferase
344 component (LgBit) and the target to the small moiety (SmBit). When both parts of the split-luciferase
345 are in close proximity, luciferase activity is restored and light is produced in the presence of furimazine-
346 based (NanoLuc) substrate[40,41] (Fig. 7A). An outstanding question in the pneumococcal cell biology
347 field is whether the muramidase MpgA and the bifunctional class A penicillin binding protein PBP1a
348 are part of the so called elongasome[42–45], the machinery responsible for cell elongation. To address this
349 question, we fused RodZ, a known component of the elongasome, as well as MpgA, to the LgBit. We
350 designed a pegRNA to insert a sequence encoding a 2-amino acid flexible linker (SG), followed by the
351 SmBit fragment (VTGYRLFEEIL), at the C-terminus of PBP1a (Fig. 7B). Of the 12 clones tested after
352 selection on aTc, 1 contained the correct fusion, signifying a selection efficiency of 8% for introduction
353 of a 39 bps-long tag, in line with the *luc*-library sequencing results showing very few clones containing
354 insertions longer than 30 bps present in the pool (Fig. 4B). After transforming the mbPE$^{S.pn}$-created
355 *pbp1A-SmBit* strain with *mpgA-LgBit* or *rodZ-LgBit*, respectively, cells were grown in microtiter plates
356 in the presence of NanoLuc luciferase substrate and bioluminescence recorded. The *pbp1A-SmBit* strain
357 was also transformed with *hu-LgBit* as a negative control since the highly expressed histone-like protein
358 HU is not expected to interact with PBP1a[40,46]. As shown in Fig. 7C, we could detect a strong and
359 reproducible signal for MpgA-LgBit and RodZ-LgBit but not for HU-LgBit, providing evidence that
360 PBP1a, MpgA and RodZ are in close proximity in space and time. To further substantiate these findings,
361 we constructed a double labelled strain in which PBP1a was tagged to a red fluorescent protein
362 (mScarlet-I-opt) on its N-terminus, under the control of its native promoter, while MpgA was tagged at
363 its N-terminus to the monomeric superfolder green fluorescent protein (GFP), also expressed from its
364 native promoter. Strain VL7454 (mScarlet-Pbp1a, GFP-MpgA) was grown in C+Y medium at 37°C
365 and mid-exponentially growing cells were imaged by epifluorescence microscopy.As shown in Fig. 7D,
366 clear co-localization between GFP-PBP1a and mScarlet-MpgA was observed, demonstrating that
367 PBP1a and MpgA are part of the same complex. Kymographs of individually GFP-tagged PBP1a and
368 MpgA also showed similar localization dynamics across the pneumococcal cell cycle (Fig. 7D). These

369 data are in line with recent work demonstrating that both PBP1a and MpgA interact with the PBP1a-
370 activating S protein and thus are likely part of the same protein complex[46].

371

372 **Sequential genome editing with mbPE$^{S.pn}$**

373 The ability to introduce several genome edits sequentially would greatly accelerate genetic engineering
374 projects in *S. pneumoniae*. To facilitate this, we constructed a second pegRNA cloning vector with
375 identical integration sites as plasmid pVL4133 but a different antibiotic resistance marker, conferring
376 resistance against chloramphenicol instead of spectinomycin (pVL7227) (Fig. 7E-F). In this way, after
377 the first genome edit has been made, the pegRNA can be replaced by transformation of the second
378 vector containing a different pegRNA. As a proof of concept, we first introduced a 63 bps deletion in
379 *luc* in strain VL6307 using pegRNA plasmid pVL6304, resulting in strain VL6926 (Δ63-*luc*). Next, we
380 designed and generated pegRNA constructs that would introduce stop mutations in either the capsule
381 gene *cps2A* or the major autolysin *lytA*. Both genes encode major pneumococcal virulence factors (Fig.
382 7E). By sequential transformation of strain VL6926 with plasmids pVL7230 (chloramphenicol
383 resistance, pegRNA-Δ*cps2A-stop*) and pVL5524 (spectinomycin resistance, pegRNA-Δ*lytA-stop*), we
384 could readily obtain strain VL7292 containing the three desired mutations (Δ63-*luc*, Δ*cps2A-stop*,
385 Δ*lytA-stop*) (Fig. 7F). Sanger sequencing verified the correct edits in two clones of the triple mutant
386 (Fig. S3A), whilst whole genome PacBio sequencing of two clones of the single mutant and two clones
387 of the triple mutant demonstrated the absence of off-target edits or genomic rearrangements (Fig. S3B
388 and S3C) (SRA accession number PRJNA1062582).

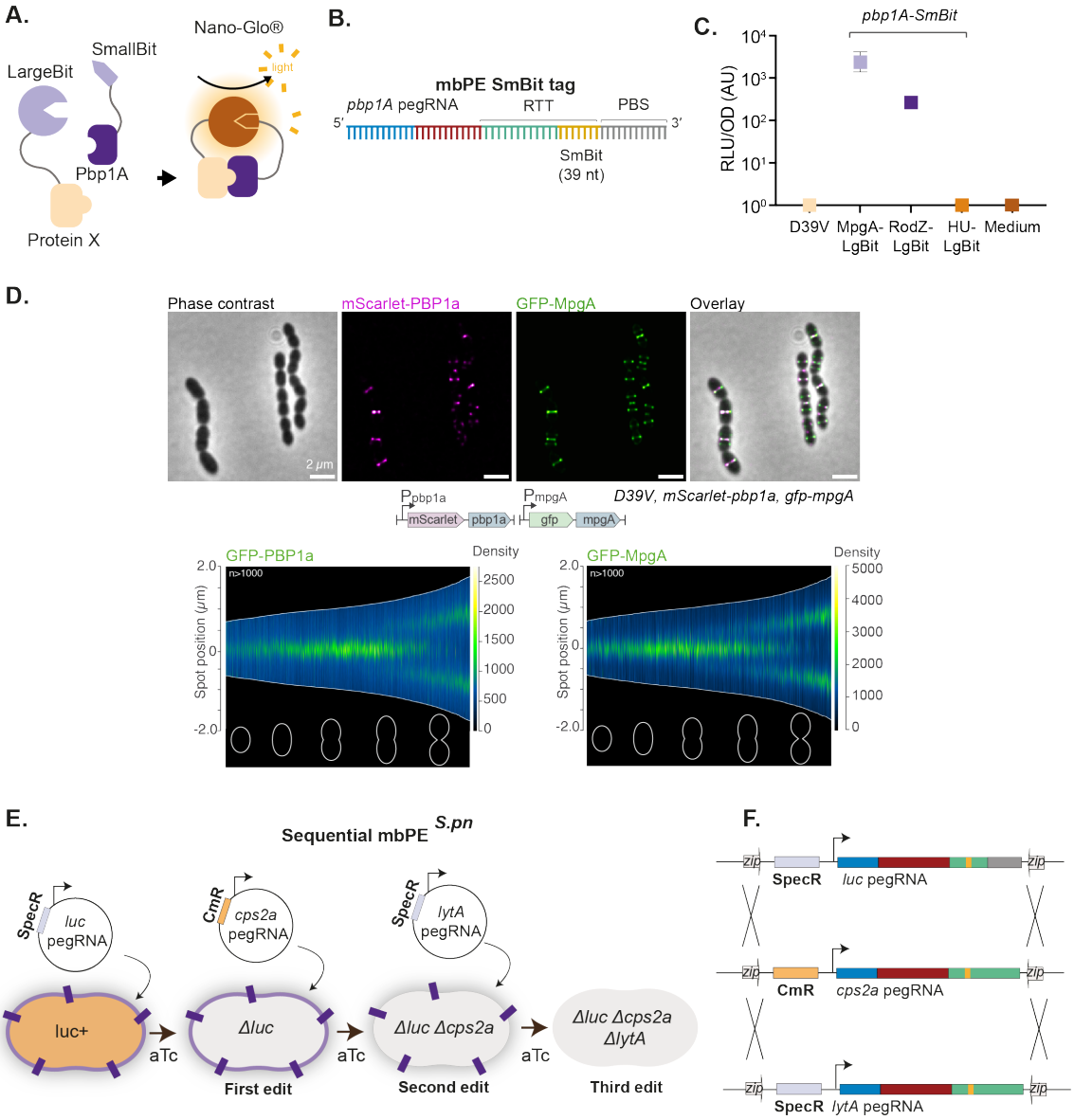


Figure 7. PBP1a is in a complex with MpgA and RodZ and sequential genome editing using mbPE. (**A**) Schematic overview of the principle of the split *luc* system. (**B**) Schematic overview of the pegRNA used to fuse a SmBit tag to *pbp1A* using mbPE. (**C**) PBP1a interacts with MpgA and RodZ. Each dot represents the average of 15 measurements of a technical replicate, with the size of the dot representing the SEM. Control strains only expressing *pbp1A-SmBit* (labeled as D39V) or *pbp1A-SmBit* together with the abundant DNA-binding protein HU-LgBit did not demonstrate bioluminescence. **(D)** PBP1a and MpgA co-localize in space and time. Live cell fluorescence microscopy of cells of strain VL7454 (mScarlet-PBP1a, GFP-MpgA) grown at 37°C in C+Y medium. Scale bar: 2 mm. Demograph of exponentially growing cells of strains VL5532 (GFP-PBP1a) and VL5816 (GFP-MpgA). (**E-F**) Sequential gene editing in *S. pneumoniae* using mbPE. Schematic overview of cloning regime with the two alternative pegRNA vector backbones (pVL4133, *spec*; pVL7227, *cat*) to generate a triple genome edited *S. pneumoniae* strain (Δ63-*luc*, Δ*cps2A-stop*, Δ*lytA-stop*) in a sequential manner.

402 # Discussion

403 The principal contribution of this work is the development of make-or-break prime editing (mbPE), a
404 highly effective bacterial genome editing system. In contrast to traditional prime editing in which a
405 nickase is used, mbPE utilizes a Cas9 nuclease fused to a reverse transcriptase. Combined with the high
406 recombinogenic potential of *S. pneumoniae*[23], the edit encoded within the pegRNA that also disrupts
407 the PAM or seed region is integrated into the genome by a mechanism that does not require RecA (Fig.
408 2E). It is tempting to speculate that, in absence of RecA, a single strand annealing mechanism[47] or
409 another kind of microhomology-mediated repair process[9], such as the RecBCD (RexAB in *S.*
410 *pneumoniae*[48]) or RecFOR loader complexes[49] play a role in consolidating the edit encoded by the
411 pegRNA during mbPE. RecJ may be a key enzyme to remove the 5′ flap and promote DNA repair (Fig.
412 2B). Since Cas9 cannot re-target correctly edited cells, as they would lack a PAM or seed region, they
413 are efficiently selected, while unedited clones die due to repeated Cas9-induced DNA breaks. This
414 model for mbPE implies that editing efficiencies might be further improved by reducing the efficiency
415 of Cas9 cleavage to allow for enough time to install the edit, something also observed for traditional
416 CRISPR gene editing in bacteria[8]. Since mbPE does not require homology-dependent recombination
417 through RecA, it might also be an efficient method in other bacteria that are less recombinogenic
418 compared to *S. pneumoniae*.

419 Recently developed nuclease-based prime editing systems in eukaryotic cells have also shown
420 improved selection efficiencies[19–21,50–52]. However, these systems often produce more undesired by-
421 products, such as insertions and deletions (indels), than the intended edits[53]. Probably because *S.*
422 *pneumoniae* does not, or very poorly, performs template-independent end joining (like most other
423 bacteria[2,9,10]), we have not observed off-target indels with mbPE$^{S.pn}$ when editing efficiencies were high.
424 Bystander mutations, if present, were rather found near the Cas9 cleavage site that are typically also
425 found during regular PE2 prime editing[7] or within mbPE$^{S.pn}$ itself (Fig. S2G). Long-read PacBio
426 sequencing of triply sequential edited strains (Fig. 7, Fig. S3B and Fig S3C) also did not show evidence
427 of genomic rearrangements or untargeted indels.

428 By screening pooled libraries containing a suite of pegRNAs with different RTT and PBS
429 lengths to introduce point mutations at a specific site (*luc*), we established that ideal PBS and RTT
430 lengths for a single adenine insertion in *luc* are 16 nts and 15 nts, respectively (Fig. 5). In addition,
431 mbPE was able to make all 12 possible nucleotide substitutions as well as insertions up to 52 bps (Figs.
432 3-4, Fig. S4). As a proof of principle, we used mbPE to tag the important amoxicillin-resistance
433 determinant PBP1A[54] with the SmBit split luciferase moiety and unequivocally showed that PBP1A is
434 in close spatial and temporal proximity to MpgA and RodZ. Deletions up to 127 bps could also be
435 efficiently introduced (Fig. 4C-E). By creating pegRNA vectors with various antibiotic resistance

436 markers, sequential genome editing is a straightforward procedure and after every consecutive edit the
437 pegRNA is simply replaced by the next (Fig. 7D-E). Together, this work demonstrates that mbPE is a
438 flexible, scalable and efficient genome editing tool for *S. pneumoniae.*

439

440 **Limitations and perspectives of mbPE**

441 Due to its natural competence, *S. pneumoniae* has been a cornerstone of molecular biology[55]. This
442 feature also enabled us to optimize mbPE for genome editing. It will now be interesting to test whether
443 mbPE is also effective in other bacteria, particularly in strains that are difficult to transform. By placing
444 both the prime editor and the pegRNA cloning cassette on a replicative plasmid with a temperature-
445 sensitive origin of replication, straightforward genome editing in hard-to-transform strains might be
446 enabled. Another advantage of *S. pneumoniae* is that its genome, which is 2.04 Mb in size in case of
447 strain D39V[35], contains 145,892 *S. pyogenes* PAM sites (NGG) distributed fairly evenly across its
448 sequence. This results in one PAM site every 14 base pairs on average. If a distance of 50 bp from a
449 PAM site is considered as the maximum distance to be efficiently edited by mbPE$^{S.pn}$ (Fig. 6B), 99.845%
450 of the D39V genome would be in range for efficient substitution editing through this system, leaving
451 only 3158 bp unavailable. Although SNPs can be efficiently introduced at locations near a PAM site,
452 the efficiency decreases as the distance from the PAM site increases (Fig. 6, Fig. S4). This is thus not a
453 major issue for *S. pneumoniae*, but could be problematic for other bacterial species, particularly those
454 with very low GC content. To obtain more genome editing flexibility, Cas9 variants with relaxed or
455 different PAM requirements[56] could be used as nuclease within the mbPE. For instance, highly
456 promising mbPE nucleases would be SpRY and SpRYc, which are engineered near-PAMless *S.*
457 *pyogenes* Cas9 variants[57,58].

458 While the *selection* efficiency of mbPE can potentially be greater than 93%, if the mutation
459 being introduced concerns a simple edit taking place at the PAM site (Fig. 2), the actual *editing*
460 efficiency is typically less than 5% (Fig. 1). This low editing efficiency thus hinders multiplex genome
461 editing to introduce multiple edits simultaneously as the chance that a cell made multiple edits before
462 being targeted by the nuclease domain is currently smaller than the selection of suppressor mutants.
463 Another limitation of prime editing are the restrictions in insertion as well as deletion lengths. Therefore,
464 future make-or-break prime editors would also benefit from improvements in reverse transcriptases and
465 pegRNA design driven by genetic screens, advances in machine learning and/or structural insights[59,60].
466 Removal of certain host factors, such as 3′→5′ DNA exonucleases and the mismatch repair system, may
467 also improve editing efficiencies[18,33,34]. Future optimized mbPEs with PAMless nucleases and efficient
468 reverse transcriptases with optimal pegRNA designs would enable multiplex editing of every base in a

469 genome. Dual pegRNA approaches may also increase the efficiency of large genomic deletions, and
470 facilitate the integration of larger DNA segments[61,62].

471 While not yet perfect, the here-described mbPE system still represents one of the most powerful
472 and versatile bacterial genome editor currently available due to its scalability, meaning the ability to
473 clone pools of pegRNAs to generate vast libraries of potential gene edits, and we foresee many exciting
474 and important applications, especially since the quality (and price) of oligo pools has significantly
475 improved in the last years. For instance, by designing a pool of pegRNAs, mbPE could be employed to
476 perform 'all codon scanning' on single genes of interest to gain new important insights into protein
477 function, protein interaction, or identify temperature-sensitive alleles. Another application could be to
478 examine gene essentiality at a genome-wide scale. Since mbPE can easily introduce mutations without
479 polar effects, large, genome-wide stop-codon or start-codon mutant libraries could be made.

480

481 **Concluding Remarks**

482 The here-described mbPE system will facilitate genome-editing approaches in the significant human
483 pathogen *S. pneumoniae* and may contribute to the discovery of new insights with regards to
484 pneumococcal cell biology as well as lead to innovative therapeutic approaches. The described methods
485 and approaches may serve as a roadmap to develop mbPE for other bacteria.

486

487 # Methods

488 **Bacterial strains, culture conditions and transformation**

489 All strains, plasmids and primers used in this study are listed in Tables S2 and S3. Pneumococcal strains
490 in this study are derivate of *S. pneumoniae* D39V[30], unless specified otherwise, and are listed in Table
491 S2. Construction of strains is described in the Supplementary information. Strains were grown in liquid
492 semi-defined C+Y medium[55] (pH 6.8) or Columbia Agar Base (CAB) supplemented with 5% sheep
493 blood at 37°C. For transformation, pneumococci were grown in C+Y medium to an optical density
494 (OD_{595nm}) of 0.1. Competence was activated by adding 100 ng/mL (CSP-1) (synthetic competence-
495 stimulating peptide 1) for 12 minutes at 37°C. The donor DNA was added to the culture and incubated
496 20 minutes at 30°C. The cultures were then diluted 10 times in C+Y and incubated at 37°C for 1.5
497 hours. Transformants were selected in CAB containing 5% blood and the appropriate antibiotic. Re-
498 streaked and confirmed clones were grown in C+Y to an OD_{595nm} of 0.3 and stored at -80°C in C+Y
499 containing 17% glycerol.

500 *E. coli* Stbl3 (Thermofisher) was used for plasmid cloning. Strains were cultured in LB or LBA
501 with appropriate antibiotics.

502

503 **Cloning of pegRNAs by Golden Gate Assembly**

504 The pPEPZ_pegRNA vector was digested with Esp3I and purified from agarose gel to ensure removal
505 of the *mCherry* fragment. The two oligonucleotides designed for the spacer, the scaffold and the
506 extension sequences of each pegRNA were annealed in TEN buffer (10 mM Tris, 1 mM EDTA, 100
507 mM NaCl, pH 8) in a thermocycler at 95°C for 5 minutes and then slowly cooled down at 25°C.
508 Equimolar ratios of each annealed pegRNA pieces were phosphorylated with T4 polynucleotide kinase
509 (New England Biolabs). The annealed and phosphorylated pieces were ligated with the
510 pPEPZ_pegRNA digested vector with T4 DNA ligase. The ligation products were transformed in *E.*
511 *coli* Stbl3 and the transformants were selected on LBA plates with the appropriate antibiotic. Details
512 about each pegRNA plasmid construction are indicated in Supplementary Information.

513

514 **Induction of PE2 and mbPE$^{S.pn}$**

515 Strains containing the PE2 nickase prime editor or the mbPE$^{S.pn}$ prime editor were grown in C+Y to an
516 OD_{595nm} of 0.3 ($3x10^8$). Serial dilutions were prepared in C+Y medium and plated in CAB plates
517 containing 5% blood and a range of aTc (between 2 - 50 ng/mL aTc) and incubated overnight at 37°C.
518 A control without aTc was included to compare the colony number with and without inducer. Single
519 colonies of plates in which approximately 10 to 50x fewer colonies were present than in the 0 aTc plates
520 at the same dilution were picked and cultured in C+Y for further experiments. Typically, when >100X
521 less colonies were present on the +aTc plates compared to the -aTc plates, editing was not efficient
522 enough compared to cutting by the nuclease and a lower aTc concentration was used for induction.

523

524 **Growth assays and luciferase activity assays**

525 Single colonies from PE2 or mbPE induction plates were picked and pre-cultured in 200 µl of C+Y
526 with no inducer and incubated at 37°C. After 2 hours of incubation, 50 µL of the colony suspension
527 were transferred to white, clear bottom 96-well plates containing 250 µL of C+Y plus 0.12 mg/mL
528 luciferin (final luciferin concentration, 0.1 mg/mL; BioSynth). For each assay D39V, uninduced
529 colonies and media were included as controls in duplicates. Absorbance (OD_{595}) and luminescence
530 (relative luminesce units, RLU) were measured for 20 hours with an interval time of 10 minutes, using
531 a Tecan (Spark) plate reader. RLU/OD values were normalized against the baseline level of medium
532 luminescence obtained for each experiment and data analyzed using GraphPad Prism.

533

534 **Nano-Glo Dual-Luciferase assay**

535 Strains were pre-cultured in C+Y to an OD_{595} of 0.3. Luminescence (relative luminesce units, RLU)
536 and Absorbance (OD_{595}) of sample triplicates supplemented with NanoGlo Live Cell Assay System

537 (Promega) as recommended by the manufacturer, were monitored using a Tecan (Spark) plate reader
538 during 15 minutes without intervals.
539
540 **pegRNA pool cloning**
541 The backbone for the cloning of the pegRNA pool (pVL4134) was linearized by XbaI digestion. The
542 linearized vector was then amplified with primers OVL7743 and OVL7744 to introduce flanking Esp3I
543 cleavage sites. Template plasmid was removed by digestion with DpnI. The ssDNA pegRNA pool
544 (Table S1, Twist Bioscience) was amplified using primers OVL7966 and OVL6481 and the PCR
545 product was excised from a 2% agarose gel using the Monarch DNA cleanup and gel extraction kit
546 (NEB). One-step Golden Gate Assembly (GGA) was used to ligate the vector and the pegRNA pool
547 using Esp3I (molar ratio vector insert 1:10). The GGA product was transformed in *E. coli* Stbl3 and
548 transformants were selected on LBA plates containing 100 µg/mL spectinomycin. After overnight
549 incubation at 37°C, several single colonies were picked to check the cloning efficiency and the rest of
550 the colonies were scraped-off from the plates to prepare glycerol stocks with the suspension. The
551 plasmid pegRNA pool was purified and transformed into *S. pneumoniae* VL5155 and transformants
552 were plated on CAB plates supplemented with 5% blood and 100 µg/mL spectinomycin and incubated
553 overnight at 37°C. The cells were scraped-off the plates and the bacterial suspension was stored at -
554 80°C in C+Y containing 17% glycerol (pool name: VL5315).
555
556 **pegRNA pool screening**
557 To determine the best conditions for the pegRNA pool screening, a pilot mbPE induction experiment
558 was performed. 100 µL of VL5355 were resuspended in 4 mL of C+Y containing 100 µg/mL
559 spectinomycin, in triplicates. Starting OD_{595} was 0.17 and the cells were grown to an OD_{595} of 0.34.
560 Serial dilutions of the bacterial suspension (up to 10^{-5}) were prepared in C+Y containing 100 µg/mL
561 spectinomycin and plated onto 75 mL CAB + 5% blood plates with and without 50 ng/mL aTc and
562 incubated overnight at 37°C. The CFU were counted, and it was determined that the selection efficiency
563 was about 10%. The dilutions to be used for the screening were chosen to be 10^{-2} for the inducer
564 condition and 10^{-3} for the no inducer condition (single colonies could be still observed and the plates
565 had a good coverage). The pegRNA pool screening was performed in quadruplicates with the same
566 conditions previously tested in the pilot experiment. All colonies grown with and without inducer were
567 scraped-off using C+Y + spectinomycin and stored with 17% glycerol at -80°C. Chromosomal DNA
568 from the pegRNA pool screening samples was extracted for Illumina library preparation using the
569 Promega kit as described previously[40].
570
571 **Targeted *luc* modification site sequencing and differential enrichment analysis**

572 Custom oligonucleotide primers (Table S3) consisting of P5 or P7 Illumina sequencing adapters,
573 Illumina sequencing barcodes (N501-504/N701-704), read 1 or read 2 sequencing primer binding sites,
574 and portions binding to the chromosomally located *luc* were used to amplify the targeted *luc*
575 modification site using Phanta high-fidelity polymerase (Vazyme). Amplified fragments were purified
576 from gel and sequenced on an Illumina MiSeq, using a 2 x 300 (600 cycles) paired-end v3 sequencing
577 kit. Illumina sequencing reads were checked for quality using FastQC (v0.11.8, available from
578 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and paired reads were merged using
579 PEAR (v0.9.6, available from https://www.h-its.org/software/pear-paired-end-read-merger/). PEAR
580 merged reads were subsequently used to count the number of occurrences of each pegRNA-encoded
581 mutation, using 2FAST2Q (v2.5.3)[39] in counter mode with consideration for the minimal Phred quality
582 score (all set to 20), and the mismatches allowed in upstream, downstream, and queried sequences (all
583 0). Counts from 2FAST2Q were imported into R, and normalized using DESeq2 (v1.40.2). DESeq2[63]
584 was subsequently used to perform the differential enrichment analysis between the control and aTc
585 induced samples. Statistical significance for differential presence was defined as an absolute
586 log2FoldChange > 1, and an adjusted p-value < 0.05. Full results of the *luc* mutation analysis can be
587 found in Table S4. Note that because of the high number of unedited, wild type *luc* alleles present in
588 the amplified pools, we were unable to accurately quantify the substitution identities and their editing
589 efficiencies for single base pair substitutions. Instead, we quantified the abundance of the pegRNAs in
590 presence and absence of aTc as a proxy for editing efficiency of single base pair substitutions.

591

592 **Sequencing of pegRNA and differential enrichment analysis**

593 Custom oligonucleotide primers (Table S3) consisting of P5 or P7 Illumina sequencing adapters,
594 Illumina sequencing barcodes (N501-504/N701-704), read 1 or read 2 sequencing primer binding sites,
595 and portions binding to the chromosomally integrated pegRNA expression were used to amplify the
596 pegRNAs using Phanta polymerase. Amplified fragments were purified and sequenced using an
597 Illumina MiniSeq instrument, with a 2 x 150 (300 cycles) paired-end mid output sequencing kit. Raw
598 Illumina sequencing reads were checked for quality using FastQC (v0.11.8, available from
599 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and paired reads were merged using
600 PEAR (v0.9.6, available from https://www.h-its.org/software/pear-paired-end-read-merger/). PEAR
601 merged reads were subsequently used to count the number of occurrences of each pegRNA-encoded
602 mutation, using 2FAST2Q (v2.5.3)[39] in counter mode with consideration for the minimal Phred quality
603 score (all set to 20), and the mismatches allowed in upstream, downstream, and queried sequences (all
604 0). Counts from 2FAST2Q were imported into R, and normalized using DESeq2 (v1.40.2). DESeq2[63]
605 was subsequently used to perform the differential enrichment analysis between the control and aTc

606 induced samples. Statistical significance for differential presence was defined as an absolute
607 log2FoldChange > 1, and an adjusted p-value < 0.05. Full results of the pegRNA abundance analysis
608 can be found in Table S5.

609

610 **Quantifications and statistical analysis**

611 Data analysis was performed using R and Prism (Graphpad). Data shown are represented as mean of at
612 least three replicates ± SEM if data came from one experiment with replicated measurement, and ± SD
613 if data came from separate experiments.

614

615 **Genome sequencing and analysis**

616 Genome sequencing was performed using PacBio HiFi sequencing on a Pacbio sequel II machine.
617 PacBio HiFi reads were assembled into circular genomes using Flye[64]. Flye-produced assemblies were
618 rotated for *dnaA* using circlator[65]. Assemblies were assessed for structural rearrangements and large
619 indels through minimap2[66], by mapping them to an *in silico* constructed reference genome containing
620 the known mutations and insertions. Genome alignments were visualized using the dotPlotly R package.
621 For SNPs and small indel calling, the raw reads from the PacBio sequencing were mapped to the
622 synthetic genomes using minimap2[66], and filtered for secondary and split alignments. Resulting
623 alignments were sorted using SAMtools, and mutations were called using bcftools, hits were filtered
624 according to previously used methods[67]. Coverage at each position was assessed using SAMtools[68].

625

626 **Data availability**

627 Sequencing data from sequencing reactions of the luciferase gene and the pegRNAs, and the PacBio
628 sequencing, are available in the Sequence Read Archive under BioProject accession code
629 PRJNA1062582.

630

631 ## Acknowledgements

632 Work in the Veening lab is supported by the Swiss National Science Foundation (SNSF) (project grants
633 310030_192517, 310030_200792 and 'AntiResist' 51NF40_180541) and ERC consolidator grant
634 771534-PneumoCaTChER. A.B.J. is supported through a Postdoctoral Fellowship grant
635 (TMPFP3_210202) from the SNSF, and a Robert Austrian Research Award from the International
636 Society of Pneumonia and Pneumococcal Diseases. X.L. was supported by Shenzhen University 2035

Program for Excellent Research (86901-00000216). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors Contributions

J.W.V. wrote the paper with input from all authors. M.R.G.., M.V.M. A.B.J., A.S.R. and J.B. performed the experiments. M.G.R., A.B.J., X.L., M.V.M., A.S.R., J.B., and J.W.V designed, analyzed and interpreted the data. All authors reviewed and approved the final version of the manuscript.

Transparency declarations

M.R.G. : None

M.V.M. : None

A.B.J. : None

A.S.R. : None

J.B. : None

X.L. : None

J.W.V. is a scientific advisory board member at i-Seq Biotechnology.

ORCID and email addresses

Name	Email address	ORCID ID
Monica Rengifo-Gonzalez	Monica.Rengifo@unil.ch	
Maria Vittoria Mazzuoli	MariaVittoria.Mazzuoli@unil.ch	0000-0002-7071-1698
Axel B. Janssen	Absalom.Janssen@unil.ch	0000-0002-9865-447X
Anne-Stephanie Rueff	Anne-Stephanie.Rueff@unil.ch	
Jessica Burnier	Jessica.Burnier@unil.ch	0009-0006-6187-8126
Xue Liu	XueLiu@szu.edu.cn	0000-0001-6485-1865
Jan-Willem Veening	Jan-Willem.Veening@unil.ch	0000-0002-3162-6634

657 # References

658 1. Ramachandran, G. & Bikard, D. Editing the microbiome the CRISPR way. *Philos. Trans. R. Soc.*
659 *B Biol. Sci.* **374**, 20180103 (2019).
660 2. Shapiro, R. S., Chavez, A. & Collins, J. J. CRISPR-based genomic tools for the manipulation of
661 genetically intractable microorganisms. *Nat. Rev. Microbiol.* **16**, 333–339 (2018).
662 3. Penewit, K. *et al.* Efficient and Scalable Precision Genome Editing in *Staphylococcus aureus*
663 through Conditional Recombineering and CRISPR/Cas9-Mediated Counterselection. *mBio* **9**,
664 e00067-18 (2018).
665 4. McAllister, K. N. & Sorg, J. A. CRISPR Genome Editing Systems in the Genus Clostridium: a
666 Timely Advancement. *J. Bacteriol.* **201**, e00219-19 (2019).
667 5. Tong, Y. *et al.* CRISPR-Cas9, CRISPRi and CRISPR-BEST-mediated genetic manipulation in
668 streptomycetes. *Nat. Protoc.* **15**, 2470–2502 (2020).
669 6. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial
670 immunity. *Science* **337**, 816–821 (2012).
671 7. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base
672 editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
673 8. Collias, D. *et al.* Systematically attenuating DNA targeting enables CRISPR-driven editing in
674 bacteria. *Nat. Commun.* **14**, 680 (2023).
675 9. Finger-Bou, M., Orsi, E., van der Oost, J. & Staals, R. H. J. CRISPR with a Happy Ending: Non-
676 Templated DNA Repair for Prokaryotic Genome Engineering. *Biotechnol. J.* **15**, 1900404 (2020).
677 10. Shuman, S. & Glickman, M. S. Bacterial DNA repair by non-homologous end joining. *Nat. Rev.*
678 *Microbiol.* **5**, 852–861 (2007).
679 11. Jiang, W. & Marraffini, L. A. CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial
680 Immunity Systems. *Annu. Rev. Microbiol.* **69**, 209–228 (2015).
681 12. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor
682 DNA. *Nature* **576**, 149–157 (2019).
683 13. Scholefield, J. & Harrison, P. T. Prime editing - an update on the field. *Gene Ther.* **28**, 396–401
684 (2021).
685 14. Hua, K., Han, P. & Zhu, J.-K. Improvement of base editors and prime editors advances precision
686 genome engineering in plants. *Plant Physiol.* **188**, 1795–1810 (2022).
687 15. Villiger, L. *et al.* CRISPR technologies for genome, epigenome and transcriptome editing. *Nat.*
688 *Rev. Mol. Cell Biol.* (2024) doi:10.1038/s41580-023-00697-6.
689 16. Zhao, Z., Shang, P., Mohanraju, P. & Geijsen, N. Prime editing: advances and therapeutic
690 applications. *Trends Biotechnol.* **41**, 1000–1012 (2023).

691 17. Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T. & Lee, S. Y. A versatile genetic engineering
692 toolkit for *E. coli* based on CRISPR-prime editing. *Nat. Commun.* **12**, 5206 (2021).
693 18. Zhang, H. *et al.* BacPE: a versatile prime-editing platform in bacteria by inhibiting DNA
694 exonucleases. *Nat. Commun.* **15**, 825 (2024).
695 19. Tao, R. *et al.* WT-PE: Prime editing with nuclease wild-type Cas9 enables versatile large-scale
696 genome editing. *Signal Transduct. Target. Ther.* **7**, 108 (2022).
697 20. Peterka, M. *et al.* Harnessing DSB repair to promote efficient homology-dependent and -
698 independent prime editing. *Nat. Commun.* **13**, 1240 (2022).
699 21. Adikusuma, F. *et al.* Optimized nickase- and nuclease-based prime editing in human and mouse
700 cells. *Nucleic Acids Res.* **49**, 10785–10795 (2021).
701 22. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an
702 engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
703 23. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial
704 genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
705 24. Synefiaridou, D. & Veening, J.-W. Harnessing CRISPR-Cas9 for Genome Editing in *Streptococcus*
706 *pneumoniae* D39V. *Appl. Environ. Microbiol.* **87**, e02762-20 (2021).
707 25. Liu, X. *et al.* Exploration of Bacterial Bottlenecks and *Streptococcus pneumoniae* Pathogenesis by
708 CRISPRi-Seq. *Cell Host Microbe* **29**, 107-120.e6 (2021).
709 26. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in
710 *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13**, 931 (2017).
711 27. Sorg, R. A., Kuipers, O. P. & Veening, J.-W. Gene expression platform for synthetic biology in the
712 human pathogen *Streptococcus pneumoniae*. *ACS Synth. Biol.* **4**, 228–239 (2015).
713 28. van Gestel, J., Hawkins, J. S., Todor, H. & Gross, C. A. Computational pipeline for designing guide
714 RNAs for mismatch-CRISPRi. *STAR Protoc.* **2**, 100521 (2021).
715 29. de Bakker, V., Liu, X., Bravo, A. M. & Veening, J.-W. CRISPRi-seq for genome-wide fitness
716 quantification in bacteria. *Nat. Protoc.* **17**, 252–281 (2022).
717 30. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the
718 endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3570–3575 (2015).
719 31. Hsu, J. Y. *et al.* PrimeDesign software for rapid and simplified design of prime editing guide RNAs.
720 *Nat. Commun.* **12**, 1034 (2021).
721 32. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**,
722 402–410 (2022).
723 33. Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing
724 outcomes. *Cell* **184**, 5635-5652.e29 (2021).
725 34. Ferreira da Silva, J. *et al.* Prime editing efficiency and fidelity are enhanced in the absence of
726 mismatch repair. *Nat. Commun.* **13**, 760 (2022).

727 35. Slager, J., Aprianto, R. & Veening, J.-W. Deep genome annotation of the opportunistic human
728 pathogen *Streptococcus pneumoniae* D39. *Nucleic Acids Res.* **46**, 9971–9989 (2018).
729 36. Martin, B., Prats, H. & Claverys, J. P. Cloning of the hexA mismatch-repair gene of *Streptococcus*
730 *pneumoniae* and identification of the product. *Gene* **34**, 293–303 (1985).
731 37. Sorg, R. A., Gallay, C., Van Maele, L., Sirard, J.-C. & Veening, J.-W. Synthetic gene-regulatory
732 networks in the opportunistic human pathogen *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci.*
733 *U. S. A.* **117**, 27608–27619 (2020).
734 38. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of *Escherichia coli*.
735 *Nucleic Acids Res.* **44**, 4243–4251 (2016).
736 39. Bravo, A. M., Typas, A. & Veening, J.-W. 2FAST2Q: a general-purpose sequence search and
737 counting program for FASTQ files. *PeerJ* **10**, e14041 (2022).
738 40. Gallay, C. *et al.* CcrZ is a pneumococcal spatiotemporal cell cycle regulator that interacts with FtsZ
739 and controls DNA replication by modulating the activity of DnaA. *Nat. Microbiol.* **6**, 1175–1187
740 (2021).
741 41. Oliveira Paiva, A. M. *et al.* The Bacterial Chromatin Protein HupA Can Remodel DNA and
742 Associates with the Nucleoid in *Clostridium difficile*. *J. Mol. Biol.* **431**, 653–672 (2019).
743 42. Perez, A. J. *et al.* Elongasome core proteins and class A PBP1a display zonal, processive movement
744 at the midcell of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2401831121
745 (2024).
746 43. Gibson, P. S. & Veening, J.-W. Gaps in the wall: understanding cell wall biology to tackle
747 amoxicillin resistance in *Streptococcus pneumoniae*. *Curr. Opin. Microbiol.* **72**, 102261 (2023).
748 44. Straume, D., Piechowiak, K. W., Kjos, M. & Håvarstein, L. S. Class A PBPs: It is time to rethink
749 traditional paradigms. *Mol. Microbiol.* **116**, 41–52 (2021).
750 45. Lenoir, C. *et al.* The morphogenic protein CopD controls the spatio-temporal dynamics of PBP1a
751 and PBP2b in *Streptococcus pneumoniae*. *mBio* **14**, e0141123 (2023).
752 46. Burnier, J. *et al.* A bacterial cell wall repair and modification system to resist host antibacterial
753 factors. *bioRxiv*. 2024.11.08.622053 Preprint at https://doi.org/10.1101/2024.11.08.622053 (2024).
754 47. Vu, T. V., Das, S., Nguyen, C. C., Kim, J. & Kim, J.-Y. Single-strand annealing: Molecular
755 mechanisms and potential applications in CRISPR-Cas-based precision genome editing.
756 *Biotechnol. J.* **17**, e2100413 (2022).
757 48. Halpern, D., Gruss, A., Claverys, J.-P. & Karoui, M. E. *rexAB* mutants in *Streptococcus*
758 *pneumoniae*. *Microbiology* **150**, 2409–2414 (2004).
759 49. Johnston, C. *et al.* RecFOR Is Not Required for Pneumococcal Transformation but Together with
760 XerS for Resolution of Chromosome Dimers Frequently Formed in the Process. *PLOS Genet.* **11**,
761 e1004934 (2015).

762 50. Li, X. *et al.* Development of a versatile nuclease prime editor with upgraded precision. *Nat.*
763 *Commun.* **14**, 305 (2023).
764 51. Jiang, T., Zhang, X.-O., Weng, Z. & Xue, W. Deletion and replacement of long genomic sequences
765 using prime editing. *Nat. Biotechnol.* **40**, 227–234 (2022).
766 52. Kweon, J. *et al.* Targeted genomic translocations and inversions generated using a paired prime
767 editing strategy. *Mol. Ther. J. Am. Soc. Gene Ther.* **31**, 249–259 (2023).
768 53. Zeng, H. *et al.* Recent advances in prime editing technologies and their promises for therapeutic
769 applications. *Curr. Opin. Biotechnol.* **86**, 103071 (2024).
770 54. Gibson, P. S., Bexkens, E., Zuber, S., Cowley, L. A. & Veening, J.-W. The acquisition of clinically
771 relevant amoxicillin resistance in *Streptococcus pneumoniae* requires ordered horizontal gene
772 transfer of four loci. *PLoS Pathog.* **18**, e1010727 (2022).
773 55. Hiller, N. L. & Orihuela, C. J. Biological puzzles solved by using *Streptococcus pneumoniae*: a
774 historical review of the pneumococcal studies that have impacted medicine and shaped molecular
775 bacteriology. *J. Bacteriol.* **0**, e00059-24 (2024).
776 56. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nat.*
777 *Commun.* **12**, 555 (2021).
778 57. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome
779 targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
780 58. Zhao, L. *et al.* PAM-flexible genome editing with an engineered chimeric Cas9. *Nat. Commun.* **14**,
781 6175 (2023).
782 59. Shuto, Y. *et al.* Structural basis for pegRNA-guided reverse transcription by a prime editor. *Nature*
783 (2024) doi:10.1038/s41586-024-07497-8.
784 60. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nat.*
785 *Rev. Genet.* **24**, 161–177 (2023).
786 61. Anzalone, A. V. *et al.* Programmable deletion, replacement, integration and inversion of large DNA
787 sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).
788 62. Wang, J. *et al.* Efficient targeted insertion of large DNA fragments without DNA donors. *Nat.*
789 *Methods* **19**, 331–340 (2022).
790 63. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
791 seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
792 64. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using
793 repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
794 65. Hunt, M. *et al.* Circlator: automated circularization of genome assemblies using long sequencing
795 reads. *Genome Biol.* **16**, 294 (2015).
796 66. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma. Oxf. Engl.* **34**, 3094–
797 3100 (2018).

67. Janssen, A. B. *et al.* Evolution of Colistin Resistance in the *Klebsiella pneumoniae* Complex Follows Multiple Evolutionary Trajectories with Variable Effects on Fitness and Virulence Characteristics. *Antimicrob. Agents Chemother.* **65**, 10.1128/aac.01958-20 (2020).
68. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).