# 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], 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 homology-directed repair of DSBs. Prime editing circumvents this need as the repair template is 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 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 targeting of *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]. By co-transforming a donor template DNA, the DSB can be repaired by homology-directed repair, resulting in the desired genome edit[1,7]. 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[2,8,9]. 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[1,10]. Especially with hard-to-transform or slow growing bacteria, each of these steps can take several days, if not weeks.

Recently, prime editors such as PE2 were described in which a Cas9 nickase is fused to an engineered reverse transcriptase[7,11]. 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[11]. Because of these advantages, prime editing has rapidly gained traction and has been successfully applied to engineer the genomes in various eukaryotes including plants[12–15]. 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[16,17].

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. 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 represents one of the most efficient and scalable bacterial genome engineering platforms currently available.

# Results

### Prime editing in *S. pneumoniae*

Several Cas9-based tools have been developed for *S. pneumoniae*[18–22], 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)[11]. 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) in *S. pneumoniae*[21]. In a first step to a functioning PE2 for *S. pneumoniae*, we mutated the alanine residue at position 10 of the $P_{tet}$-*dcas9* construct back to an aspartic acid to create $P_{tet}$-*cas9*$^{H840A}$. The engineered MMLVRT domain and a linker domain from PE2 were codon optimized for optimal

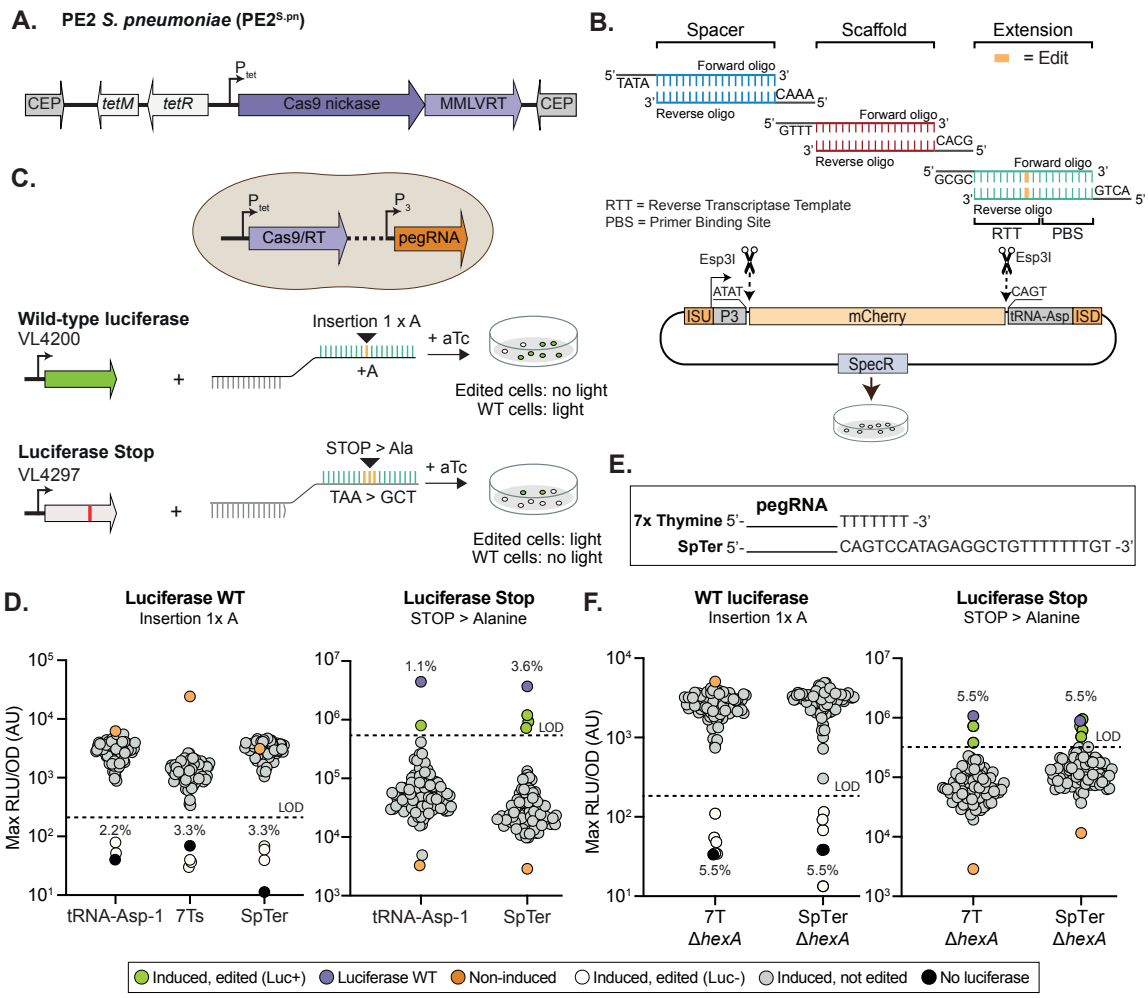
83 expression in low GC-rich bacteria such as *S. pneumoniae* and cloned to the C-terminus of the Cas9
84 nickase resulting in strain VL4199 ($P_{tet}$-*cas9*$^{H840A}$-MMLVRT or $P_{tet}$-PE2$^{S.pn}$) (Fig. 1A).

85 Next, a pegRNA construction and *S. pneumoniae* integration vector was designed in which the
86 pegRNA is constitutively expressed from the constitutive P3 promoter including its +1 adenine
87 transcription start site to ensure that all cloned pegRNAs are efficiently transcribed and do not contain
88 undesirable uridines at the 5' end[22–24]. For easy cloning, an *mCherry* cassette flanked by Esp3I
89 (isoschizomer of BsmBI) sites is present allowing for rapid, oligo-based, Golden Gate cloning of
90 pegRNAs (Fig. 1B). Sequence elements facilitating one-step library preparation for NGS were also
91 introduced[25] resulting in pegRNA cloning vectors pVL4132, pVL4133 and pVL4393 (carrying 7
92 thymine nucleotides, a *S. pneumoniae* terminator or a *S. pneumoniae tRNA-asp-1* after the extension
93 sequence, respectively). pegRNAs consist of four elements: 1) a 20 nt spacer sequence for targeting of
94 the prime editor, 2) a scaffold sequence for binding to Cas9, 3) a primer binding site (PBS) that allows
95 the 3′-end of the nicked DNA strand to hybridize to the pegRNA, and 4) a reverse transcriptase template
96 (RTT) containing the desired edit[11]. Since bacterial transcription termination leaves nucleotides at the
97 3'end of the pegRNA, which may interfere with efficient prime editing downstream, we cloned *S.*
98 *pneumoniae tRNA-asp-1* directly following the pegRNA (Fig. 1B). This strategy has been previously
99 used to produce multiple individual sgRNAs from a single polycistronic operon as the endogenous
100 tRNA-processing system precisely cleaves both ends of the *tRNA-asp-1* precursor[26]. To test whether
101 the PE2$^{S.pn}$ is functional, we designed several pegRNAs using PrimeDesign[27] to introduce either a
102 premature stop codon in the firefly luciferase gene (*luc*) expressed in pneumococcal strain VL4200, or
103 to introduce a repairing mutation in strain VL4297 in which *luc* has a premature stop codon (Fig. 1C).
104 Interestingly, induction of PE2$^{S.pn}$ together with a targeting pegRNA did not result in significant growth
105 delay, suggesting that *S. pneumoniae* can tolerate and readily repair ssDNA breaks (Fig. S1). As shown
106 in Fig. 1D, approximately 2% of clones in which PE2$^{S.pn}$ was induced demonstrated the desired
107 phenotype. Sanger sequencing confirmed the bioluminescence observations (Fig. S1B).

108 Recently, it was shown that a structured 3'-end of a pegRNA increases stability and improves
109 editing efficiency in human cells[28]. In addition, a functional mismatch repair system was shown to
110 negatively affect editing efficiencies[29,30]. To test whether the addition of a structured RNA motif at the
111 3'-end would also improve PE2 editing efficiency in *S. pneumoniae*, we replaced the tRNA sequence
112 for the structured and highly efficient *S. pneumoniae rpsT* transcriptional terminator[31] (SpTer) or a
113 stretch of 7 thymine nucleotides (7 Ts) (Fig. 1E). As shown in Fig. 1D, both pegRNAs showed improved
114 efficiency of inserting a single A in *luc* (~3%) through PE2$^{S.pn}$-mediated editing, compared to the tRNA-
115 Asp1 system. Repairing an internal stop codon (TAA to GCT) in *luc* was approximately 3 times more
116 efficient with the SpTer terminator compared to the tRNA-Asp1 construct (Fig. 1D). Deleting the
117 pneumococcal mismatch repair gene *hexA*[32] further enhanced PE2$^{S.pn}$ editing efficiency in the constructs

118 with either the 7T or SpTer terminators (Fig. 1F). These results demonstrate that prime editing using
119 PE2$^{S.pn}$ in combination with a structured pegRNA works in *S. pneumoniae* and can be used to
120 incorporate insertions and mutations, albeit at a relatively low selection efficiency.

121




122

123 **Figure 1. Prime editing 2 (PE2) in *S. pneumoniae.*** (**A**) Schematic overview of PE2$^{S.pn}$. (**B**) Schematic overview
124 of the general cloning strategy of pegRNAs (shown is the *tRNA-asp-1* containing vector pVL4393 as backbone).
125 (**C**) Editing strategy used to test efficiency of PE2$^{S.pn}$ in pneumococcal strains expressing either a functional
126 luciferase (VL4200) or a *luc* gene interrupted by a premature stop codon (VL4297). (**D**) Induction of PE2$^{S.pn}$ by
127 aTc in combination with expression of a pegRNA without 3' leader, by processing of the *tRNA-asp-1*, leads to
128 accurate gene editing in approximately 2% of *S. pneumoniae* cells. Induction of PE2$^{S.pn}$ in combination with 3'
129 extended pegRNAs provides an editing efficiency of approximately 3%. Cells were grown in C+Y medium
130 containing luciferin substrate. The maximal bioluminescence (relative light units; RLU) reached, normalized by
131 the optical density (OD), of each clone is shown (AU, arbitrary units). RLU/OD values were normalized against
132 the baseline level of medium luminescence obtained for each experiment. (**E**) Schematic overview of pegRNAs

133 containing the *S. pneumoniae rpsT* terminator (SpTer) or a stretch of 7 Ts. (**F**) Prime editing is slightly enhanced
134 in the absence of mismatch repair in *S. pneumoniae*.
135

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

137 Since prime editing depends on successful hybridization of the 3'end of the nicked DNA to the
138 pegRNA[11], we wondered if prime editing would also be functional when, instead of a nick, a DSB is
139 introduced. Recent work in human HEK293T cells showed efficient prime editing utilizing Cas9-
140 induced DSBs that largely depend on NHEJ[33–35]. To test if DSB-driven prime editing (here after called
141 make-or-break prime editing: mbPE) would also work in bacteria lacking NHEJ, we first mutated the
142 alanine residue at position 840 within the Cas9 nickase moiety of PE2$^{S.pn}$ back to a histidine. This led
143 to full restoration of wild type Cas9, resulting in strain VL5149 ($P_{tet}$-*cas9*-MMLVRT or mbPE$^{S.pn}$) (Fig.
144 2A). This construct was combined with *luc* expression (strain VL5155) and the same set of pegRNAs
145 containing the SpTer or 7 Ts as previously used to verify the efficiency of the nickase-based PE2$^{S.pn}$
146 (Fig. 1E). All pegRNAs were designed to either disrupt the PAM or the seed region within *luc*. Thus,
147 only correctly edited bacteria are expected to survive the expression of mbPE$^{S.pn}$, as unedited repairs
148 will lead to continued DSBs, hence the name make-or-break prime editing. Induction of mbPE$^{S.pn}$ with
149 aTc, in the presence of a targeting pegRNA, caused severe growth defects indicating that DSBs were
150 formed (Fig. S2A). Strikingly, surviving colonies (approximately 1 - 5%, depending on the pegRNA),
151 predominantly displayed the correct luciferase phenotype (Fig. 2B). Sanger sequencing confirmed that
152 over 93% of colonies had the precise edit as encoded by the pegRNA template (Fig. S2B). Colonies
153 that did not display the correct *luc*-edit either had suppressor mutations in the Cas9 domain, or deletions
154 in the $P_{tet}$ promoter driving the expression of mbPE$^{S.pn}$ (Fig. S2C). Note that, if desired, strains can
155 easily be cured from the prime editing system by transforming with plasmid pPEPY to exchange the
156 mbPE$^{S.pn}$ cassette with a kanamycin resistance marker[36]. PacBio HiFi long read whole genome
157 sequencing of 2 edited clones did not reveal signs of off-target editing nor any genomic rearrangements
158 (SRA accession number PRJNA1062582).
159

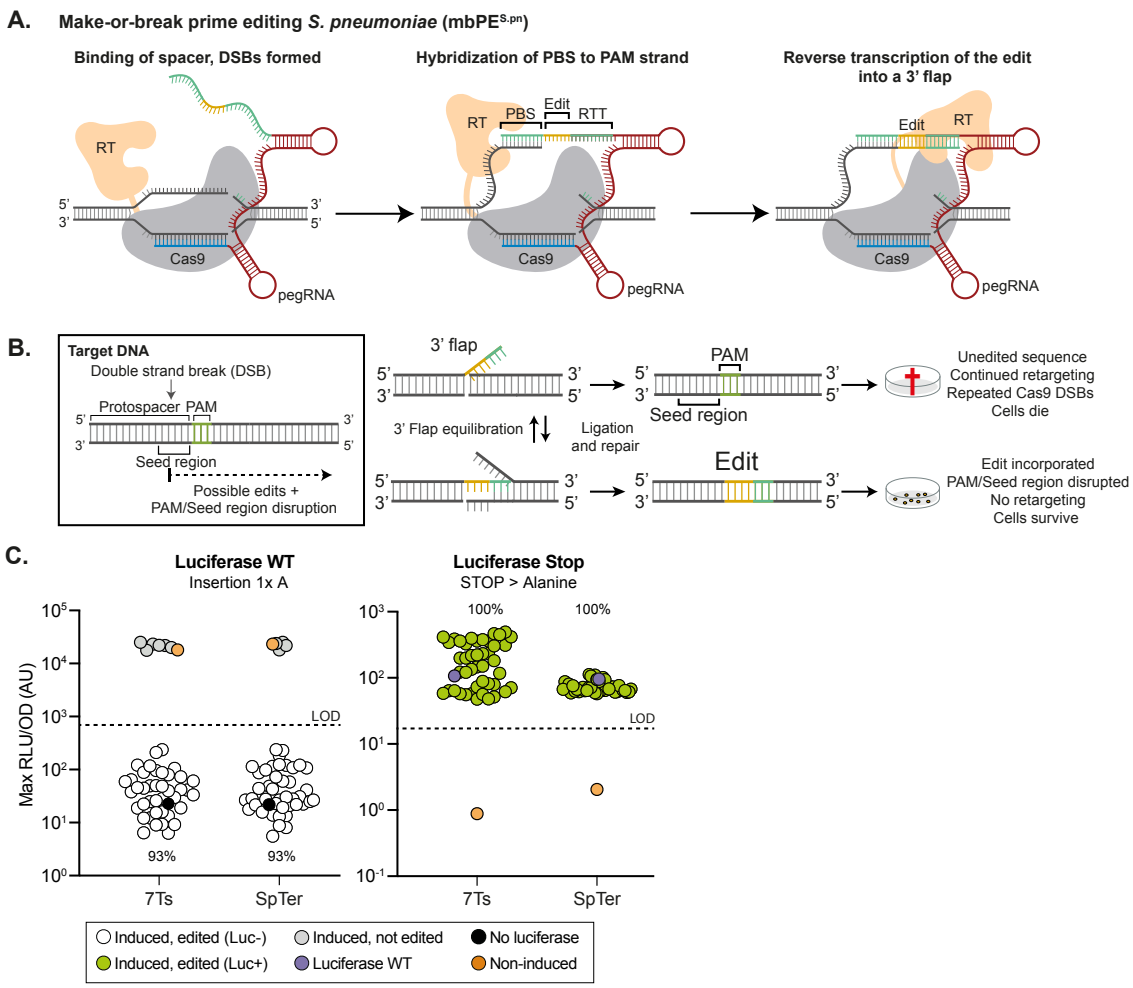



**Figure 2. Make-or-break prime editing selectively enriches correctly edited clones.** (**A**) Schematic overview of mbPE$^{S.pn}$. Binding of the spacer sequence to the target region leads to formation of DSBs by the Cas9 domain. After, the PBS can hybridize to its complementary sequence. The reverse transcriptase domain of mbPE$^{S.pn}$ subsequently synthesizes cDNA using the RTT as template. (**B**) Schematic depicting the mechanism leading to selective survival of edited clones. The hybridization of the 3'-flap containing the edit is in equilibrium with the unedited strand. Repair with the edited region will prevent repeated DSBs by Cas9. (**C**) Induction of mbPE$^{S.pn}$ by aTc in combination with expression of a structured pegRNA leads to accurate gene editing. Approximately 93% of surviving *S. pneumoniae* cells after induction are accurately edited. Cells were grown in C+Y medium containing luciferin substrate. The maximal bioluminescence reached of each clone is shown (RLU: relative light units. OD: optical density).

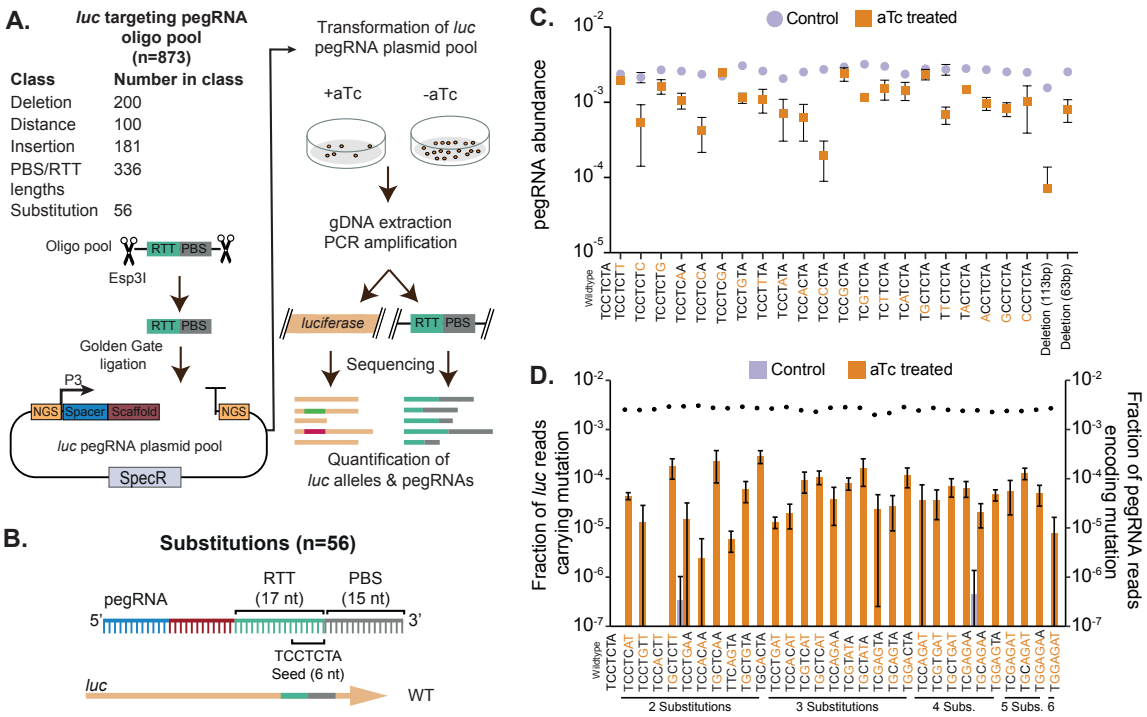
## Scalable pooled pegRNA library construction shows efficient installment of base pair substitutions

To explore the potential of mbPE$^{S.pn}$, we generated a pool of different pegRNAs targeting *luc* (Table S1), followed by Golden Gate cloning (Fig. 3A). A total of 873 oligos were designed and synthesized as a pool, containing different PBS and RTT sequences to introduce various genome edits. The pegRNA pool was cloned into plasmid pVL4134 that already contains the *luc*-targeting spacer sequence. The

178 resulting pool was subsequently used to transform strain VL5155 carrying *luc* and the aTc-inducible
179 mbPE$^{S.pn}$ (see Methods). The library was plated on agar lacking or containing 20 ng/ml aTc to induce
180 mbPE$^{S.pn}$. Quadruplicate induced and uninduced experiments, each with a minimum of 40000 colonies,
181 were pooled, and their pegRNA contents and edited *luc* region were sequenced (Fig. 3A). In the
182 uninduced pool, all but 12 of the designed pegRNAs were present, and we did not observe some of the
183 pegRNAs encoding the longest insertions. Strikingly, most pegRNAs were also present in the induced
184 pool albeit at varying abundances. By counting and comparing the number of reads coming of each
185 pegRNA in both conditions, we could estimate an editing efficiency score for each pegRNA for a
186 specific mutation (Fig. 3A). By also sequencing the *luc* gene, we confirmed that most of the designed
187 edits were correctly made in the pool of mutants (see Methods).

188 First, we examined whether mbPE$^{S.pn}$ was able to install various substitutions (a>t, a>c, a>g,
189 t>a, t>c, t>g, c>a, c>t, c>g) of targets within a seed region. We designed pegRNAs that have the same
190 20 nt *luc* spacer and 15 nt PBS, but varying RTTs encoding the different substitutions (and the necessary
191 *luc* seed disruptions) (Fig. 3B, Table S1). As shown in Fig. 3C, pegRNAs encoding all substitutions
192 were present at high frequencies under induced conditions suggesting that the installment of single base
193 pair substitutions is highly efficient with mbPE$^{S.pn}$. Next, we wondered if mbPE$^{S.pn}$ could introduce
194 multiple substitutions simultaneously. Indeed, all tested substitutions, up to the 6 adjacent substitutions
195 tested (CCTCTA> GGAGAT), could readily be introduced (Fig. 3D).

196




197

198 **Figure 3. mbPE$^{S.pn}$ is highly scalable and efficient in making base pair substitutions. (A)** Cloning strategy to
199 produce a library of pegRNAs targeting *luc* from a ssDNA oligo pool. The oligo pool was amplified by PCR as
200 well as the backbone of pVL4134 that includes the *luc* spacer sequence and assembled by Golden Gate
201 digestion/ligation followed by transformation to *S. pneumoniae* containing *luc* and mbPE2$^{S.pn}$. Chromosomal
202 DNA was isolated from colonies grown in absence or presence of aTc (to induce mbPE2$^{S.pn}$) and PCR amplified
203 using staggered primers followed by illumina sequencing (see Methods). 2FAST2Q[37] was used to extract and
204 count pegRNAs and *luc* alleles. (**B**) Schematic overview of the pegRNA design encoding for different base pair
205 substitutions. (**C-D**) Single base pair substitutions and multiple base pair substitutions are installed with high
206 selection efficiency. (**C**) pegRNA abundance of +/- aTc samples are shown. (**D**) The fraction of *luc* reads carrying
207 the desired mutation is shown for induced and non-induced samples on the left Y-axis. The average fraction of
208 pegRNA reads for the substitutions is shown as dots on the right Y-axis.
209

210 **Efficient DNA insertion and deletion with mbPE**

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

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

232

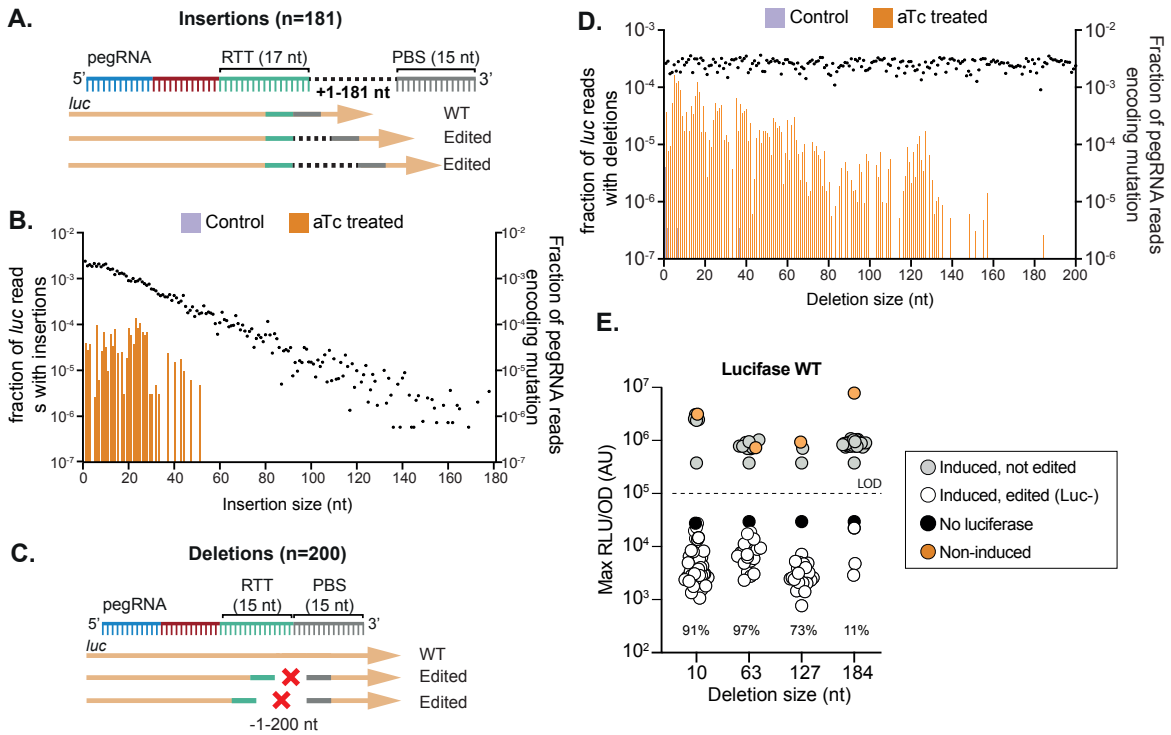



**Figure 4. mbPE for DNA insertions and genome deletions.** (**A**, **C**) Schematic overview of the approach used to test insertion and deletion efficiency in *luc* using mbPE. For pegRNAs encoding insertions, additional nucleotides were added between the RTT and PBS. For deletions, the RTT portion 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 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.

## 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[11]. 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* seed sequence, 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 (Fig. 5B). Together, these experiments provide guidelines for optimal pegRNA design and shows that

mbPE is a versatile editing tool capable of introducing single point mutations, multiple simultaneous substitutions, insertions and deletions.

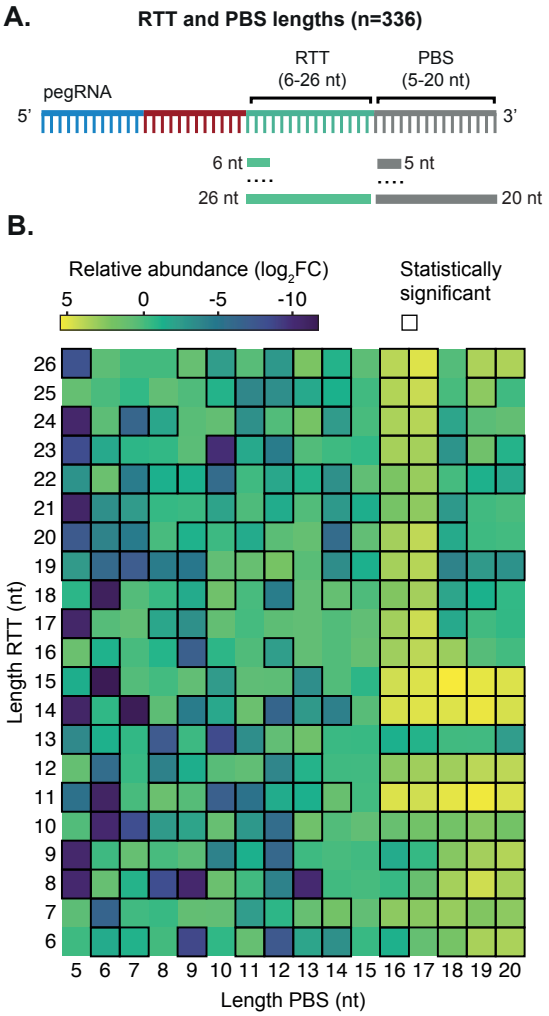



**Figure 5. pegRNA design rules for mbPE$^{S.pn}$.** (**A**) Schematic overview of the approach used to test mbPE$^{S.pn}$ efficiency at inserting a single adenine in *luc*. (**B**) Combinations of RTT and PBS with statistically significant differences in pegRNA abundance between uninduced vs induced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

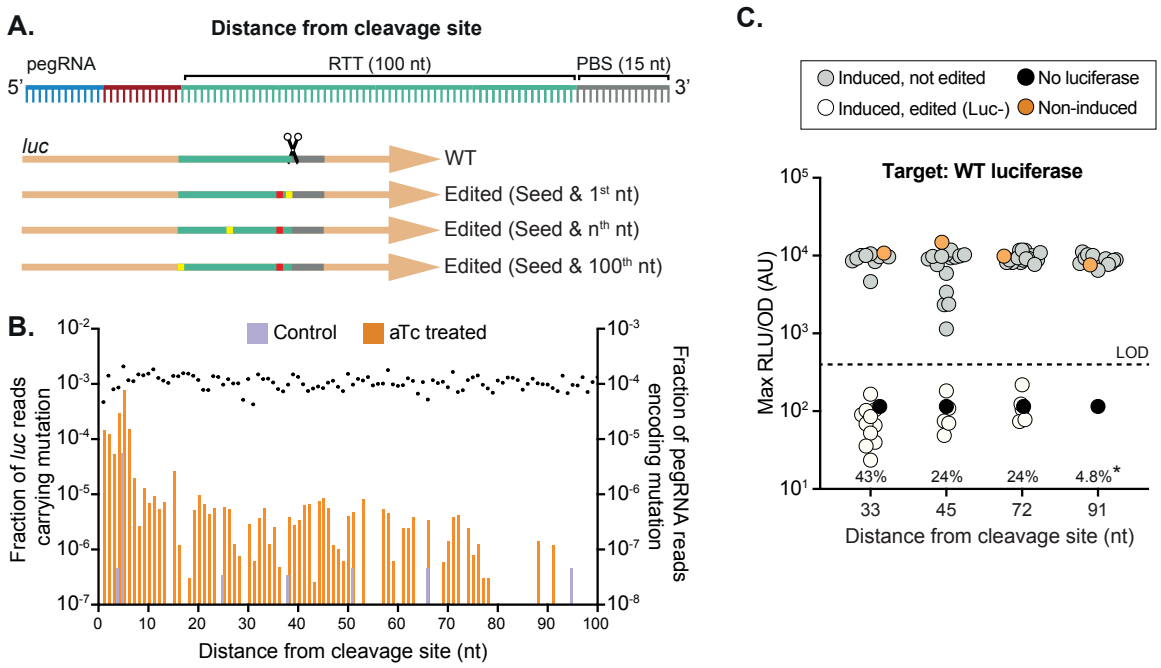
## Exploring the editing distance from the seed region

The strategies employed in this study guide the Cas9 nuclease moiety of mbPE$^{S.pn}$ to cut within the seed region near the PAM motif as dictated by the spacer sequence, while the reverse transcriptase (RT) component of mbPE$^{S.pn}$ introduces the desired genome edit. This approach prevents Cas9 from retargeting the same location. The dependency on a PAM site is, however, a possible limitation of the mbPE system. For mbPE to be broadly applicable for the introduction of mutations, even in genomic

271 regions without a nearby PAM sequence, editing should also be possible at sites distal from the PAM.
272 To determine the maximum distance from the Cas9 cut site at which a single nucleotide substitution is
273 possible, we designed a set of 100 pegRNAs with identical spacers. Each pegRNA contained RTT
274 sequences to mutate the seed region (preventing retargeting) while introducing a SNP at varying
275 distances from the PAM site (1-100 nts from the restriction site; Fig. 6A). As shown in Fig. 6B, edits
276 were observed as far as 91 bp from the PAM site.

277 Several pegRNAs were individually cloned, each with a mutation that destroys the seed region
278 to insert a silent mutation in *luc* and a substitution further from the PAM site. This showed that SNPs
279 could be efficiently introduced at a distance of 33 base pairs from the PAM (Fig. 6C). Editing was also
280 possible at further distances, but with increasingly reduced efficiency, similar to what was observed in
281 the pooled screen (Fig. 6B). Together, this demonstrates that mbPE$^{S.pn}$ is highly versatile and can, in
282 principle, be used to edit nearly every base in the pneumococcal genome (see Discussion).

283




284

285 **Figure 6. mbPE can introduce edits at sites distal from the PAM.** (**A**) Schematic overview of the RTT design
286 strategy used to edit bases distal from the PAM site. (**B**) Fraction of correctly edited *luc* alleles and pegRNA read
287 counts of DNA extracted from cells grown in the absence (control) or presence of aTc. Edits are observed up to
288 91 bp from the PAM site. (**C**) Validation of the maximum editing distance through individually cloned pegRNAs.
289 The mutations conferred by the pegRNAs encoding a mutation at distances of 33, and 45 bps from the PAM site,
290 confer a premature stop codon, whilst the mutations conferred by the pegRNAs encoding mutation either 72 or
291 91 bps from the PAM site resulted in an glycine to cysteine (72 bps) or isoleucine to asparagine (91 bps)
292 substitution. The percentages indicate the fraction of clones that were phenotypically observed to be edited. The
293 mutation at 91 nts (*) could only be confirmed by sequencing individual clones, as the amino acid change does

294 not affect luciferase function. This showed an editing efficiency of 4.8 % (1 out of 21 tested colonies) with the
295 correct genome edit.
296

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

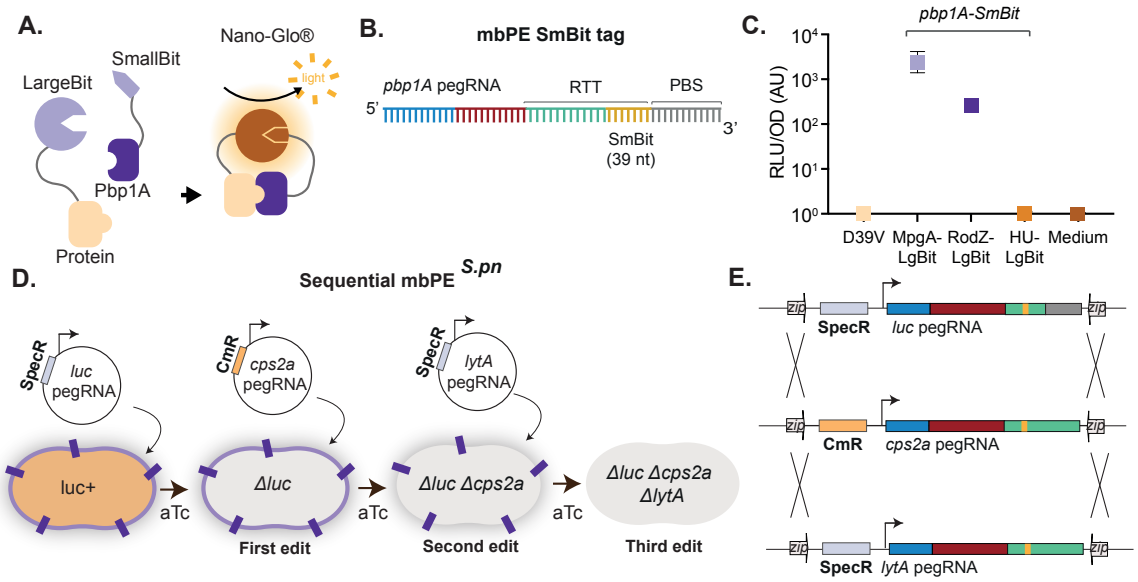
298 Our pooled pegRNA assays indicated that DNA up to 52 bps can be inserted into the pneumococcal
299 genome using mbPE (Fig. 4B). To test whether mbPE could be used to introduce protein tags, we
300 considered the split luciferase tag. The use of split luciferase has emerged as a powerful approach to
301 identify protein-protein interactions in live bacterial cells in which the bait is fused to the large luciferase
302 component (LgBit) and the target to the small moiety (SmBit). When both parts of the split-luciferase
303 are in close proximity, luciferase activity is restored and light is produced in the presence of furimazine-
304 based (NanoLuc) substrate[38,39] (Fig. 7A). An outstanding question in the pneumococcal cell biology
305 field is whether the muramidase MpgA and the bifunctional class A penicillin binding protein PBP1a
306 are part of the so called elongasome[40–43], the machinery responsible for cell elongation. To address this
307 question, we fused RodZ, a known component of the elongasome, as well as MpgA, to the LgBit. We
308 designed a pegRNA to insert a sequence encoding a 2-amino acid flexible linker (SG), followed by the
309 SmBit fragment (VTGYRLFEEIL), at the C-terminus of PBP1a (Fig. 7B). Of the 12 clones tested after
310 selection on aTc, 1 contained the correct fusion, signifying a selection efficiency of 8% for introduction
311 of a 39 bps-long tag, in line with the pooled library sequencing results (Fig. 4B). After transforming the
312 mbPE$^{S.pn}$-created *pbp1A-SmBit* strain with *mpgA-LgBit* or *rodZ-LgBit*, respectively, cells were grown
313 in microtiter plates in the presence of NanoLuc luciferase substrate and bioluminescence recorded. As
314 shown in Fig. 7C, we could detect a strong and reproducible signal in both strains, providing evidence
315 that PBP1a, MpgA and RodZ are in close proximity in space and time in live pneumococcal cells.

316

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

318 The ability to introduce several genome edits sequentially would greatly accelerate genetic engineering
319 projects in *S. pneumoniae*. To facilitate this, we constructed a second pegRNA cloning vector with
320 identical integration sites as plasmid pVL4133 but a different antibiotic resistance marker, conferring
321 resistance against chloramphenicol instead of spectinomycin (pVL7227) (Fig. 7D). In this way, after
322 the first genome edit has been made, the pegRNA can be replaced by transformation of the second
323 vector containing a different pegRNA. As a proof of concept, we first introduced a 63 bps deletion in
324 *luc* in strain VL6307 using pegRNA plasmid pVL6304, resulting in strain VL6926 (Δ63-*luc*). Next, we
325 designed and generated pegRNA constructs that would introduce stop mutations in either the capsule
326 gene *cps2A* or the major autolysin *lytA*. Both genes encode major pneumococcal virulence factors (Fig.
327 7D). By sequential transformation of strain VL6926 with plasmids pVL7230 (chloramphenicol
328 resistance, pegRNA- Δ*cps2A-stop*) and pVL5524 (spectinomycin resistance, Δ*lytA-stop*), we could

329 readily obtain strain VL7292 containing the three desired mutations (Δ63-*luc*, Δ*cps2A-stop*, Δ*lytA-*
330 *stop*). Whole genome PacBio resequencing of two of the triple mutants verified the correct edits and
331 demonstrated the absence of off-target edits or genomic rearrangements (SRA accession number
332 PRJNA1062582).
333



334

335 **Figure 7. PBP1a is in a complex with MpgA and RodZ and sequential genome editing using mbPE.** (**A**)
336 Schematic overview of the principle of the split *luc* system. (**B**) Schematic overview of the pegRNA used to fuse
337 a SmBit tag to *pbp1A* using mbPE. (**C**) PBP1a interacts with MpgA and RodZ. Each dot represents the average
338 of 15 measurements of a technical replicate, with the size of the dot representing the SEM. Control strains only
339 expressing *pbp1A-SmBit* (D39V) or *pbp1A-SmBit* together with the abundant DNA-binding protein HU-LgBit did
340 not demonstrate bioluminescence. (**D-E**) Sequential gene editing in *S. pneumoniae* using mbPE. Schematic
341 overview of cloning regime with the two alternative pegRNA vector backbones (pVL4133, *spec*; pVL7227, *cat*)
342 to generate a triple genome edited *S. pneumoniae* strain (Δ63-*luc*, Δ*cps2A-stop*, Δ*lytA-stop*) in a sequential
343 manner.
344

345 # Discussion

346 The principal contribution of this work is the development of make-or-break prime editing (mbPE), a
347 highly effective bacterial genome editing system. In contrast to traditional prime editing in which a
348 nickase is used, mbPE utilizes a Cas9 nuclease fused to a reverse transcriptase. Combined with the high
349 recombinogenic potential of *S. pneumoniae*[19], Cas9-induced DSBs are efficiently repaired using the edit
350 encoded within the pegRNA that also disrupts the PAM or seed region. Since Cas9 cannot re-target
351 correctly edited cells, as they would lack a PAM or seed region, they are efficiently selected, while
352 unedited clones die due to repeated Cas9-induced DNA breaks. Recently developed nuclease-based
353 prime editing systems in eukaryotic cells have also shown improved selection efficiencies[33–35,44–46].
354 However, these systems often produce more undesired by-products, such as insertions and deletions
355 (indels), than the intended edits[47]. Probably because *S. pneumoniae* does not, or very poorly, performs
356 template-independent end joining (like most other bacteria[2,8,9]), we have not observed off-target indels
357 with mbPE$^{S.pn}$ when editing efficiencies were high. Bystander mutations, if present, were rather found
358 near the Cas9 cleavage site that are typically also found during regular PE2 prime editing[7] or within
359 mbPE$^{S.pn}$ itself (Fig.S2C). Long-read PacBio sequencing of triply sequential edited strains (Fig. 7) also
360 did not show evidence of genomic rearrangements or untargeted indels.

361 By screening pooled libraries containing a suite of pegRNAs with different RTT and PBS
362 lengths to introduce point mutations, we established that ideal PBS and RTT lengths for a single adenine
363 insertion are 16 nts and 15 nts, respectively (Fig. 5). In addition, mbPE was able to make all tested
364 nucleotide substitutions as well as insertions up to 52 bps (Figs. 3-4). As a proof of principle, we used
365 mbPE to tag the important amoxicillin-resistance determinant Pbp1A[48] with the SmBit split luciferase
366 moiety and unequivocally showed that Pbp1A is in close spatial and temporal proximity to MpgA and
367 RodZ. Deletions up to 127 bps could also be efficiently introduced (Fig. 4C-E). By creating pegRNA
368 vectors with various antibiotic resistance markers, sequential genome editing is a straightforward
369 procedure and after every consecutive edit the pegRNA is simply replaced by the next (Fig. 7D-E).
370 Together, this work demonstrates that mbPE is a flexible, scalable and efficient genome editing tool for
371 *S. pneumoniae.*

372

373 **Limitations and perspectives of mbPE**

374 Due to its natural competence, *S. pneumoniae* has been a cornerstone of molecular biology[49]. This
375 feature also enabled us to optimize mbPE for genome editing. It will now be interesting to test whether
376 mbPE is also effective in other bacteria, particularly in strains that are difficult to transform. Another
377 advantage of *S. pneumoniae* is that its genome, which is 2.04 Mb in size in case of strain D39V[31],

contains 145,892 *S. pyogenes* PAM sites (NGG) distributed fairly evenly across its sequence. This results in one PAM site every 14 base pairs on average. If a distance of 50 bp from a PAM site is considered as the maximum distance to be efficiently edited by mbPE$^{S.pn}$ (Fig. 6B), 99.845% of the D39V genome would be in range for efficient substitution editing through this system, leaving only 3158 bp unavailable. Although SNPs can be efficiently introduced at locations near a PAM site, the efficiency decreases as the distance from the PAM site increases (Fig. 6). This is thus not a major issue for *S. pneumoniae*, but could be problematic for other bacterial species, particularly those with very low GC content. To obtain more genome editing flexibility, Cas9 variants with relaxed or different PAM requirements[50] could be used as nuclease within the mbPE. For instance, highly promising mbPE nucleases would be SpRY and SpRYc, which are engineered near-PAMless *S. pyogenes* Cas9 variants[51,52].

While the *selection* efficiency of mbPE is greater than 93%, if the mutation being introduced concerns a simple edit taking place near the PAM site (Fig. 2), the actual *editing* efficiency is typically less than 5% (Fig. 1). This low editing efficiency thus hinders multiplex genome editing to introduce multiple edits simultaneously. Another limitation of prime editing are the restrictions in insertion as well as deletion lengths. Therefore, future make-or-break prime editors would also benefit from improvements in reverse transcriptases and pegRNA design driven by genetic screens, advances in machine learning and/or structural insights[53,54]. Future optimized mbPEs with PAMless nucleases and efficient reverse transcriptases with optimal pegRNA designs would enable multiplex editing of every base in a genome, increase the efficiency of large genomic deletions, and facilitate the integration of larger DNA segments.

While not yet perfect, the here-described mbPE system still represents one of the most powerful and versatile bacterial genome editor currently available due to its scalability, and we foresee many exciting and important applications, especially since the quality (and price) of oligo pools has significantly improved in the last years. For instance, by designing a pool of pegRNAs, mbPE could be employed to perform 'all codon scanning' on single genes of interest to gain new important insights into protein function, protein interaction, or identify temperature-sensitive alleles. Another application could be to examine gene essentiality at a genome-wide scale. Since mbPE can easily introduce mutations without polar effects, large, genome-wide stop-codon or start-codon mutant libraries could be made. Testing such libraries, with and without mbPE inducer under various environmental conditions, followed by pegRNA sequencing and quantification of their abundance, would give precise genome-wide fitness estimates in a single experiment. This could provide insights into gene essentiality without the downsides of polar effects that methods such as Tn-Seq and CRISPRi-seq suffer from[25,55]. Alternatively, inducible promoters could be inserted by mbPE in front of every gene in a genome using a pool of pegRNAs to test gene fitness in presence or absence of inducer. Another approach using

413 pooled pegRNAs would be to encode protein-tags such as the split-luc tag here described (Fig. 7A-C)
414 or any short immunogenic epitope (e.g. FLAG, HA, 6xHIS). pegRNAs could be designed to tag every
415 gene in a bacterial genome. This will, for instance, allow testing of protein-protein interactions at the
416 genome-wide scale or the preparation of bait pools in protein-purification strategies. To reduce the
417 number of false positives, wild type Cas9 or a second copy of the mbPE can be simultaneously
418 expressed with the prime editor to further enhance selection of correctly edited clones and reduce the
419 selection of mutants in the editor itself.

420

421 **Concluding Remarks**

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

426

# 427 Methods

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

429 All strains, plasmids and primers used in this study are listed in Tables S2 and S3. Pneumococcal strains
430 in this study are derivate of *S. pneumoniae* D39V[30], unless specified otherwise, and are listed in Table
431 S2. Construction of strains is described in the Supplementary information. Strains were grown in liquid
432 semi-defined C+Y medium[55] (pH 6.8) or Columbia Agar Base (CAB) supplemented with 5% sheep
433 blood at 37°C. For transformation, pneumococci were grown in C+Y medium to an optical density
434 ($OD_{595nm}$) of 0.1. Competence was activated by adding 100 ng/mL (CSP-1) (synthetic competence-
435 stimulating peptide 1) for 12 minutes at 37°C. The donor DNA was added to the culture and incubated
436 20 minutes at 30°C. The cultures were then diluted 10 times in C+Y and incubated at 37°C for 1.5
437 hours. Transformants were selected in CAB containing 5% blood and the appropriate antibiotic. Re-
438 streaked and confirmed clones were grown in C+Y to an $OD_{595nm}$ of 0.3 and stored at -80°C in C+Y
439 containing 17% glycerol.

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

442

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

444 The pPEPZ_pegRNA vector was digested with Esp3I and purified from agarose gel to ensure removal
445 of the *mCherry* fragment. The two oligonucleotides designed for the spacer, the scaffold and the

446 extension sequences of each pegRNA were annealed in TEN buffer (10 mM Tris, 1 mM EDTA, 100
447 mM NaCl, pH 8) in a thermocycler at 95°C for 5 minutes and then slowly cooled down at 25°C.
448 Equimolar ratios of each annealed pegRNA pieces were phosphorylated with T4 polynucleotide kinase
449 (New England Biolabs). The annealed and phosphorylated pieces were ligated with the
450 pPEPZ_pegRNA digested vector with T4 DNA ligase. The ligation products were transformed in *E.*
451 *coli* Stbl3 and the transformants were selected on LBA plates with the appropriate antibiotic. Details
452 about each pegRNA plasmid construction are indicated in Supplementary Information.
453
454 **Induction of PE2 and mbPE$^{S.pn}$**
455 Strains containing the PE2 nickase prime editor or the mbPE$^{S.pn}$ prime editor were grown in C+Y to an
456 $OD_{595nm}$ of 0.3 ($3x10^8$). Serial dilutions were prepared in C+Y medium and plated in CAB plates
457 containing 5% blood and 50 ng/mL aTc and incubated overnight at 37°C. A control without aTc was
458 included for the $1x10^{-4}$ dilution to compare the colony number with and without inducer. Single colonies
459 were picked and cultured in C+Y for further experiments.
460
461 **Growth assays and luciferase activity assays**
462 Single colonies from PE2 or mbPE induction plates were picked and pre-cultured in 200 µl of C+Y
463 with no inducer and incubated at 37°C. After 2 hours of incubation, 50 µL of the colony suspension
464 were transferred to white, clear bottom 96-well plates containing 250 µL of C+Y plus 0.12 mg/mL
465 luciferin (final luciferin concentration, 0.1 mg/mL; BioSynth). For each assay D39V, uninduced
466 colonies and media were included as controls in duplicates. Absorbance ($OD_{595}$) and luminescence
467 (relative luminesce units, RLU) were measured for 20 hours with an interval time of 10 minutes, using
468 a Tecan (Spark) plate reader. RLU/OD values were normalized against the baseline level of medium
469 luminescence obtained for each experiment and data analyzed using GraphPad Prism.
470
471 **Nano-Glo Dual-Luciferase assay**
472 Strains were pre-cultured in C+Y to an $OD_{595}$ of 0.3. Luminescence (relative luminesce units, RLU)
473 and Absorbance ($OD_{595}$) of sample triplicates supplemented with NanoGlo Live Cell Assay System
474 (Promega) as recommended by the manufacturer, were monitored using a Tecan (Spark) plate reader
475 during 15 minutes without intervals.
476
477 **pegRNA pool cloning**
478 The backbone for the cloning of the pegRNA pool (pVL4134) was linearized by XbaI digestion. The
479 linearized vector was then amplified with primers OVL7743 and OVL7744 to introduce flanking Esp3I
480 cleavage sites. Template plasmid was removed by digestion with DpnI. The ssDNA pegRNA pool
481 (Table S1, Twist Bioscience) was amplified using primers OVL7966 and OVL6481 and the PCR

482 product was excised from a 2% agarose gel using the Monarch DNA cleanup and gel extraction kit
483 (NEB). One-step Golden Gate Assembly (GGA) was used to ligate the vector and the pegRNA pool
484 using Esp3I (molar ratio vector insert 1:10). The GGA product was transformed in *E. coli* Stbl3 and
485 transformants were selected on LBA plates containing 100 µg/mL spectinomycin. After overnight
486 incubation at 37°C, several single colonies were picked to check the cloning efficiency and the rest of
487 the colonies were scraped-off from the plates to prepare glycerol stocks with the suspension. The
488 plasmid pegRNA pool was purified and transformed into *S. pneumoniae* VL5155 and transformants
489 were plated on CAB plates supplemented with 5% blood and 100 µg/mL spectinomycin and incubated
490 overnight at 37°C. The cells were scraped-off the plates and the bacterial suspension was stored at -
491 80°C in C+Y containing 17% glycerol (pool name: VL5315).
492
493 **pegRNA pool screening**
494 To determine the best conditions for the pegRNA pool screening, a pilot mbPE induction experiment
495 was performed. 100 µL of VL5355 were resuspended in 4 mL of C+Y containing 100 µg/mL
496 spectinomycin, in triplicates. Starting $OD_{595}$ was 0.17 and the cells were grown to an $OD_{595}$ of 0.34.
497 Serial dilutions of the bacterial suspension (up to $10^{-5}$) were prepared in C+Y containing 100 µg/mL
498 spectinomycin and plated onto 75 mL CAB + 5% blood plates with and without 50 ng/mL aTc and
499 incubated overnight at 37°C. The CFU were counted, and it was determined that the selection efficiency
500 was about 10%. The dilutions to be used for the screening were chosen to be $10^{-2}$ for the inducer
501 condition and $10^{-3}$ for the no inducer condition (single colonies could be still observed and the plates
502 had a good coverage). The pegRNA pool screening was performed in quadruplicates with the same
503 conditions previously tested in the pilot experiment. All colonies grown with and without inducer were
504 scraped-off using C+Y + spectinomycin and stored with 17% glycerol at -80°C. Chromosomal DNA
505 from the pegRNA pool screening samples was extracted for Illumina library preparation using the
506 Promega kit as described previously[38].
507
508 **Targeted *luc* modification site sequencing and differential enrichment analysis**
509 Custom oligonucleotide primers (Table S3) consisting of P5 or P7 Illumina sequencing adapters,
510 Illumina sequencing barcodes (N501-504/N701-704), read 1 or read 2 sequencing primer binding sites,
511 and portions binding to the chromosomally located *luc* were used to amplify the targeted *luc*
512 modification site using Phanta high-fidelity polymerase (Vazyme). Amplified fragments were purified
513 from gel and sequenced on an Illumina MiSeq, using a 2 x 300 (600 cycles) paired-end v3 sequencing
514 kit. Illumina sequencing reads were checked for quality using FastQC (v0.11.8, available from
515 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and paired reads were merged using
516 PEAR (v0.9.6, available from https://www.h-its.org/software/pear-paired-end-read-merger/). PEAR

517 merged reads were subsequently used to count the number of occurrences of each pegRNA-encoded
518 mutation, using 2FAST2Q (v2.5.3)[37] in counter mode with consideration for the minimal Phred quality
519 score (all set to 20), and the mismatches allowed in upstream, downstream, and queried sequences (all
520 0). Counts from 2FAST2Q were imported into R, and normalized using DESeq2 (v1.40.2). DESeq2[56]
521 was subsequently used to perform the differential enrichment analysis between the control and aTc
522 induced samples. Statistical significance for differential presence was defined as an absolute
523 log2FoldChange > 1, and an adjusted p-value < 0.05. Full results of the *luc* mutation analysis can be
524 found in Table S4. Note that because of the high number of unedited, wild type *luc* alleles present in
525 the amplified pools, we were unable to accurately quantify the substitution identities and their editing
526 efficiencies for single base pair substitutions. Instead, we quantified the abundance of the pegRNAs in
527 presence and absence of aTc as a proxy for editing efficiency of single base pair substitutions.

528

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

530 Custom oligonucleotide primers (Table S3) consisting of P5 or P7 Illumina sequencing adapters,
531 Illumina sequencing barcodes (N501-504/N701-704), read 1 or read 2 sequencing primer binding sites,
532 and portions binding to the chromosomally integrated pegRNA expression were used to amplify the
533 pegRNAs using Phanta polymerase. Amplified fragments were purified and sequenced using an
534 Illumina MiniSeq instrument, with a 2 x 150 (300 cycles) paired-end mid output sequencing kit. Raw
535 Illumina sequencing reads were checked for quality using FastQC (v0.11.8, available from
536 https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and paired reads were merged using
537 PEAR (v0.9.6, available from https://www.h-its.org/software/pear-paired-end-read-merger/). PEAR
538 merged reads were subsequently used to count the number of occurrences of each pegRNA-encoded
539 mutation, using 2FAST2Q (v2.5.3)[37] in counter mode with consideration for the minimal Phred quality
540 score (all set to 20), and the mismatches allowed in upstream, downstream, and queried sequences (all
541 0). Counts from 2FAST2Q were imported into R, and normalized using DESeq2 (v1.40.2). DESeq2[56]
542 was subsequently used to perform the differential enrichment analysis between the control and aTc
543 induced samples. Statistical significance for differential presence was defined as an absolute
544 log2FoldChange > 1, and an adjusted p-value < 0.05. Full results of the pegRNA abundance analysis
545 can be found in Table S5.

546

547 **Quantifications and statistical analysis**

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

551

552 **Genome sequencing and analysis**

553 Genome sequencing was performed using PacBio HiFi sequencing on a Pacbio sequel II machine.
554 PacBio HiFi reads were assembled into circular genomes using Flye[57]. Flye-produced assemblies were
555 rotated for *dnaA* using circlator[58]. Assemblies were assessed for structural rearrangements and large
556 indels through progressiveMauve[59], by mapping them to an *in silico* constructed reference genome
557 containing the known mutations and insertions. For SNPs and small indel calling, the raw reads from
558 the PacBio sequencing were mapped to the synthetic genomes using minimap2[60], and filtered for
559 secondary and split alignments. Resulting alignments were sorted using SAMtools, and mutations were
560 called using bcftools. Coverage at each position was assessed using SAMtools[61].

561

562 **Data availability**

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

566

567 ## Acknowledgements

568 Work in the Veening lab is supported by the Swiss National Science Foundation (SNSF) (project grants
569 310030_192517, 310030_200792 and 'AntiResist' 51NF40_180541) and ERC consolidator grant
570 771534-PneumoCaTChER. A.B.J. is supported through a Postdoctoral Fellowship grant
571 (TMPFP3_210202) from the SNSF. X.L. was supported by Shenzhen University 2035 Program for
572 Excellent Research (86901-00000216). The funders had no role in study design, data collection and
573 analysis, decision to publish, or preparation of the manuscript.

574

575 ## Authors Contributions

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

579

580 ## Transparency declarations

581 M.R.G.: None

582 M.V.M.: None

583 A.B.J. : None

584 A.S.R. : None

585 X.L. : None

586 J.W.V.: None

587

588 ## 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 | |
| Xue Liu | XueLiu@szu.edu.cn | 0000-0001-6485-1865 |
| Jan-Willem Veening | Jan-Willem.Veening@unil.ch | 0000-0002-3162-6634 |


589

590

591 ## References


592 1. Ramachandran, G. & Bikard, D. Editing the microbiome the CRISPR way. *Philos. Trans. R. Soc.*
593 *B Biol. Sci.* **374**, 20180103 (2019).

594 2. Shapiro, R. S., Chavez, A. & Collins, J. J. CRISPR-based genomic tools for the manipulation of
595 genetically intractable microorganisms. *Nat. Rev. Microbiol.* **16**, 333–339 (2018).

596 3. Penewit, K. *et al.* Efficient and Scalable Precision Genome Editing in Staphylococcus aureus
597 through Conditional Recombineering and CRISPR/Cas9-Mediated Counterselection. *mBio* **9**,
598 e00067-18 (2018).

599 4. McAllister, K. N. & Sorg, J. A. CRISPR Genome Editing Systems in the Genus Clostridium: a
600 Timely Advancement. *J. Bacteriol.* **201**, e00219-19 (2019).

601 5. Tong, Y. *et al.* CRISPR-Cas9, CRISPRi and CRISPR-BEST-mediated genetic manipulation in
602 streptomycetes. *Nat. Protoc.* **15**, 2470–2502 (2020).

603 6. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial
604 immunity. *Science* **337**, 816–821 (2012).
605 7. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base
606 editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
607 8. Finger-Bou, M., Orsi, E., van der Oost, J. & Staals, R. H. J. CRISPR with a Happy Ending: Non-
608 Templated DNA Repair for Prokaryotic Genome Engineering. *Biotechnol. J.* **15**, 1900404 (2020).
609 9. Shuman, S. & Glickman, M. S. Bacterial DNA repair by non-homologous end joining. *Nat. Rev.*
610 *Microbiol.* **5**, 852–861 (2007).
611 10. Jiang, W. & Marraffini, L. A. CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial
612 Immunity Systems. *Annu. Rev. Microbiol.* **69**, 209–228 (2015).
613 11. Anzalone, A. V. *et al.* Search-and-replace genome editing without double-strand breaks or donor
614 DNA. *Nature* **576**, 149–157 (2019).
615 12. Scholefield, J. & Harrison, P. T. Prime editing - an update on the field. *Gene Ther.* **28**, 396–401
616 (2021).
617 13. Hua, K., Han, P. & Zhu, J.-K. Improvement of base editors and prime editors advances precision
618 genome engineering in plants. *Plant Physiol.* **188**, 1795–1810 (2022).
619 14. Villiger, L. *et al.* CRISPR technologies for genome, epigenome and transcriptome editing. *Nat.*
620 *Rev. Mol. Cell Biol.* (2024) doi:10.1038/s41580-023-00697-6.
621 15. Zhao, Z., Shang, P., Mohanraju, P. & Geijsen, N. Prime editing: advances and therapeutic
622 applications. *Trends Biotechnol.* **41**, 1000–1012 (2023).
623 16. Tong, Y., Jørgensen, T. S., Whitford, C. M., Weber, T. & Lee, S. Y. A versatile genetic engineering
624 toolkit for E. coli based on CRISPR-prime editing. *Nat. Commun.* **12**, 5206 (2021).
625 17. Zhang, H. *et al.* BacPE: a versatile prime-editing platform in bacteria by inhibiting DNA
626 exonucleases. *Nat. Commun.* **15**, 825 (2024).
627 18. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an
628 engineered CRISPR-Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
629 19. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial
630 genomes using CRISPR-Cas systems. *Nat. Biotechnol.* **31**, 233–239 (2013).
631 20. Synefiaridou, D. & Veening, J.-W. Harnessing CRISPR-Cas9 for Genome Editing in Streptococcus
632 pneumoniae D39V. *Appl. Environ. Microbiol.* **87**, e02762-20 (2021).
633 21. Liu, X. *et al.* Exploration of Bacterial Bottlenecks and Streptococcus pneumoniae Pathogenesis by
634 CRISPRi-Seq. *Cell Host Microbe* **29**, 107-120.e6 (2021).
635 22. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in
636 Streptococcus pneumoniae. *Mol. Syst. Biol.* **13**, 931 (2017).
637 23. Sorg, R. A., Kuipers, O. P. & Veening, J.-W. Gene expression platform for synthetic biology in the
638 human pathogen Streptococcus pneumoniae. *ACS Synth. Biol.* **4**, 228–239 (2015).

639 24. van Gestel, J., Hawkins, J. S., Todor, H. & Gross, C. A. Computational pipeline for designing guide
640 RNAs for mismatch-CRISPRi. *STAR Protoc.* **2**, 100521 (2021).
641 25. de Bakker, V., Liu, X., Bravo, A. M. & Veening, J.-W. CRISPRi-seq for genome-wide fitness
642 quantification in bacteria. *Nat. Protoc.* **17**, 252–281 (2022).
643 26. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the
644 endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 3570–3575 (2015).
645 27. Hsu, J. Y. *et al.* PrimeDesign software for rapid and simplified design of prime editing guide RNAs.
646 *Nat. Commun.* **12**, 1034 (2021).
647 28. Nelson, J. W. *et al.* Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**,
648 402–410 (2022).
649 29. Chen, P. J. *et al.* Enhanced prime editing systems by manipulating cellular determinants of editing
650 outcomes. *Cell* **184**, 5635-5652.e29 (2021).
651 30. Ferreira da Silva, J. *et al.* Prime editing efficiency and fidelity are enhanced in the absence of
652 mismatch repair. *Nat. Commun.* **13**, 760 (2022).
653 31. Slager, J., Aprianto, R. & Veening, J.-W. Deep genome annotation of the opportunistic human
654 pathogen Streptococcus pneumoniae D39. *Nucleic Acids Res.* **46**, 9971–9989 (2018).
655 32. Martin, B., Prats, H. & Claverys, J. P. Cloning of the hexA mismatch-repair gene of Streptococcus
656 pneumoniae and identification of the product. *Gene* **34**, 293–303 (1985).
657 33. Tao, R. *et al.* WT-PE: Prime editing with nuclease wild-type Cas9 enables versatile large-scale
658 genome editing. *Signal Transduct. Target. Ther.* **7**, 108 (2022).
659 34. Peterka, M. *et al.* Harnessing DSB repair to promote efficient homology-dependent and -
660 independent prime editing. *Nat. Commun.* **13**, 1240 (2022).
661 35. Adikusuma, F. *et al.* Optimized nickase- and nuclease-based prime editing in human and mouse
662 cells. *Nucleic Acids Res.* **49**, 10785–10795 (2021).
663 36. Keller, L. E., Rueff, A.-S., Kurushima, J. & Veening, J.-W. Three New Integration Vectors and
664 Fluorescent Proteins for Use in the Opportunistic Human Pathogen Streptococcus pneumoniae.
665 *Genes* **10**, (2019).
666 37. Bravo, A. M., Typas, A. & Veening, J.-W. 2FAST2Q: a general-purpose sequence search and
667 counting program for FASTQ files. *PeerJ* **10**, e14041 (2022).
668 38. Gallay, C. *et al.* CcrZ is a pneumococcal spatiotemporal cell cycle regulator that interacts with FtsZ
669 and controls DNA replication by modulating the activity of DnaA. *Nat. Microbiol.* **6**, 1175–1187
670 (2021).
671 39. Oliveira Paiva, A. M. *et al.* The Bacterial Chromatin Protein HupA Can Remodel DNA and
672 Associates with the Nucleoid in Clostridium difficile. *J. Mol. Biol.* **431**, 653–672 (2019).

673 40. Perez, A. J. *et al.* Elongasome core proteins and class A PBP1a display zonal, processive movement
674 at the midcell of Streptococcus pneumoniae. *BioRxiv Prepr. Serv. Biol.* 2024.01.10.575112 (2024)
675 doi:10.1101/2024.01.10.575112.
676 41. Gibson, P. S. & Veening, J.-W. Gaps in the wall: understanding cell wall biology to tackle
677 amoxicillin resistance in Streptococcus pneumoniae. *Curr. Opin. Microbiol.* **72**, 102261 (2023).
678 42. Straume, D., Piechowiak, K. W., Kjos, M. & Håvarstein, L. S. Class A PBPs: It is time to rethink
679 traditional paradigms. *Mol. Microbiol.* **116**, 41–52 (2021).
680 43. Lenoir, C. *et al.* The morphogenic protein CopD controls the spatio-temporal dynamics of PBP1a
681 and PBP2b in Streptococcus pneumoniae. *mBio* **14**, e0141123 (2023).
682 44. Li, X. *et al.* Development of a versatile nuclease prime editor with upgraded precision. *Nat.*
683 *Commun.* **14**, 305 (2023).
684 45. Jiang, T., Zhang, X.-O., Weng, Z. & Xue, W. Deletion and replacement of long genomic sequences
685 using prime editing. *Nat. Biotechnol.* **40**, 227–234 (2022).
686 46. Kweon, J. *et al.* Targeted genomic translocations and inversions generated using a paired prime
687 editing strategy. *Mol. Ther. J. Am. Soc. Gene Ther.* **31**, 249–259 (2023).
688 47. Zeng, H. *et al.* Recent advances in prime editing technologies and their promises for therapeutic
689 applications. *Curr. Opin. Biotechnol.* **86**, 103071 (2024).
690 48. Gibson, P. S., Bexkens, E., Zuber, S., Cowley, L. A. & Veening, J.-W. The acquisition of clinically
691 relevant amoxicillin resistance in Streptococcus pneumoniae requires ordered horizontal gene
692 transfer of four loci. *PLoS Pathog.* **18**, e1010727 (2022).
693 49. Hiller, N. L. & Orihuela, C. J. Biological puzzles solved by using Streptococcus pneumoniae: a
694 historical review of the pneumococcal studies that have impacted medicine and shaped molecular
695 bacteriology. *J. Bacteriol.* **0**, e00059-24 (2024).
696 50. Collias, D. & Beisel, C. L. CRISPR technologies and the search for the PAM-free nuclease. *Nat.*
697 *Commun.* **12**, 555 (2021).
698 51. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome
699 targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).
700 52. Zhao, L. *et al.* PAM-flexible genome editing with an engineered chimeric Cas9. *Nat. Commun.* **14**,
701 6175 (2023).
702 53. Shuto, Y. *et al.* Structural basis for pegRNA-guided reverse transcription by a prime editor. *Nature*
703 (2024) doi:10.1038/s41586-024-07497-8.
704 54. Chen, P. J. & Liu, D. R. Prime editing for precise and highly versatile genome manipulation. *Nat.*
705 *Rev. Genet.* **24**, 161–177 (2023).
706 55. Cain, A. K. *et al.* A decade of advances in transposon-insertion sequencing. *Nat. Rev. Genet.* **21**,
707 526–540 (2020).

708 56. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-
709 seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
710 57. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using
711 repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
712 58. Hunt, M. *et al.* Circlator: automated circularization of genome assemblies using long sequencing
713 reads. *Genome Biol.* **16**, 294 (2015).
714 59. Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene
715 gain, loss and rearrangement. *PloS One* **5**, e11147 (2010).
716 60. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinforma. Oxf. Engl.* **34**, 3094–
717 3100 (2018).
718 61. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–
719 2079 (2009).
720

721