Prime Editing Efficiency and Fidelity are Enhanced in the Absence of Mismatch Repair

Ferreira da Silva J[1,2#], Oliveira GP[1#], Arasa-Verge EA[1], Kagiou C[1,2], Moretton A[1,2], Timelthaler G[1], Jiricny J[3], Loizou JI[1,2*]

[1]Institute of Cancer Research, Department of Medicine I, Comprehensive Cancer Centre, Medical University of Vienna, Borschkegasse 8a, 1090 Vienna, Austria

[2]CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

[3]Institute of Biochemistry of the ETH Zurich, Otto-Stern-Weg 3, 8093 Zurich, Switzerland

[#]Equal contribution

*Correspondence

Correspondence: joanna.loizou@meduniwien.ac.at (JIL)

Running title

Enhancement of Prime Editing Efficiency and Fidelity by Ablation of Mismatch Repair

Keywords

Prime editing; mismatch repair; genome engineering; DNA repair

Abstract

Prime editing (PE) is a powerful genome engineering approach that enables the introduction of base substitutions, insertions and deletions into any given genomic locus. However, the efficiency of PE varies widely and depends not only on the genomic region targeted, but also on the genetic background of the edited cell. To determine which cellular factors affect PE efficiency, we carried out a focused genetic screen targeting 32 DNA repair factors, spanning all reported repair pathways. We show that, depending on cell line and type of edit, ablation of mismatch repair (MMR) affords a 2-17 fold increase in PE efficiency, across several human cell lines, types of edits and genomic loci. The accumulation of the key MMR factors MLH1 and MSH2 at PE sites argues for direct involvement of MMR in PE control. Our results shed new light on the mechanism of PE and suggest how its efficiency might be optimised.

Introduction

CRISPR-Cas9-based genome editing technologies are powerful new tools of functional genomics, with considerable potential as future therapeutics (Jinek et al., 2012). However, the efficiency of currently available genome editing protocols is limited. Moreover, the process gives rise to undesirable side products that hinder the implementation of this technology in clinical settings. To overcome these hurdles, there is need to identify the DNA metabolic pathways and molecular mechanisms that govern editing outcomes, as well as the activities of these pathways in different cellular and tissue contexts (Richardson et al., 2018; Yeh et al., 2019; Ferreira da Silva et al., 2021; Hussmann et al., 2021). The first generation of Cas9-based genome engineering tools used nucleases that could be directed to any desired region of the genome by a single-guide RNA (sgRNA). Following the targeting of a site-specific DNA double-strand break (DSB), the endogenous DNA end-joining pathways frequently repair this lesion in an error prone manner, leading to insertions or deletions (indels) that give rise to loss-of-function alleles (Bothmer et al., 2017). This approach was further adapted to include either a single- or double-stranded donor template containing the desired edit. Here, the DSB is processed by homology-directed repair (HDR), which catalyses the insertion of the donor template that includes the edit. Unlike the former approach, which generates random indels, the latter method permits the introduction of desired indels, as well as point mutations, into the genome (Yeh et al., 2019). However, since HDR is inefficient, depends on potentially deleterious DSBs and requires cell division, alternative approaches were needed.

Amongst such alternative approaches are Base Editing (BE) and Prime Editing (PE). The former uses nucleobase modification chemistry to efficiently and precisely incorporate single nucleotide variants into the genome of cells (Komor et al., 2016; Gaudelli et al., 2017; Gu et al., 2021), but its scope is limited to single-base substitutions. This led to the development of PE as a highly versatile genome editing approach that allows for the targeted insertion of indels, point mutations and combinations thereof into the genome (Anzalone et al., 2019). PE utilises a fusion of a Cas9(H840A) nickase (Jinek et al., 2012) and reverse transcriptase (RT) that is targeted to a precise genomic region by a PE guide RNA (pegRNA). The pegRNA includes the desired sequence change, as well as a short 3’ terminal extension complementary to the 5’ sequence upstream from the nick within the target site. Annealing of the 3’ terminus of the pegRNA to the 3’ segment of the nicked DNA strand generates a substrate for the RT, which copies the RNA template and thus incorporates the desired edit into the 3’ extension of the nick. Dissociation of the RNA and annealing of the DNA strands generates a 3’ flap containing the edit. Transient melting and reannealing of the nicked target site give rise to a mixture of molecules containing either 3’ or 5’ flaps. Successful installation of the desired edit requires removal of the 5’ flap and ligation of the resulting nick to yield a DNA heteroduplex

containing the edit in the RT-synthesised strand. The editing outcome of this method, referred to as PE2, depends on the resolution of this heteroduplex. Utilising an additional sgRNA that directs nicking to the original DNA strand, either concurrently to the edit installation (PE3), or subsequently (PE3b), increases PE efficiency (Anzalone et al., 2019). The increased efficiency in PE3 strategies has been suggested to require the DNA repair pathway known as DNA Mismatch Repair (MMR) that would function in the repair of the nicked, non-edited strand, by utilising the edited strand as template (Petri et al., 2021; Scholefield and Harrison, 2021).

Due to its versatility, PE has been used in a wide variety of models, such as zebrafish (Petri et al., 2021), rice and wheat (Lin et al., 2020), mouse (Liu et al., 2020) and human stem cells (Sürün et al., 2020). A notable feature of PE is its highly variable rates across different genetic backgrounds, even within the same genomic locus and using the same pegRNA (Anzalone et al., 2019). To address whether this could be explained by different DNA repair capacities, we performed a targeted genetic screen aimed at identifying DNA repair factors involved in PE. Here, we uncover an inhibitory role for MMR pathway in PE and show that MMR proteins localise to sites of PE to directly counteract edit installation, rather than promote it. Thus, deletion or transient depletion of MMR factors increase PE efficiency and fidelity across different edit sites, types and cell lines.

Results

A targeted genetic screen identifies the DNA repair pathway mismatch repair as inhibitory for prime editing

To investigate the DNA repair requirements for PE, we conducted a targeted genetic screen, utilising a collection of isogenic knockouts in the human near-haploid HAP1 cell line (**Supplementary Data 1**). The 32 targeted genes were selected to represent divergent functions within all known human DNA repair pathways. The library thus provided a comprehensive coverage of the DNA damage response. The cell lines received the PE machinery, including the Cas9(H840A)-RT and a pegRNA encoding a 5-base pair (bp) deletion in the *HEK3* locus. PE efficiency was determined by amplicon sequencing of the genomic locus.

Wild-type HAP1 cells were remarkably inefficient at PE (<1% alleles edited). In contrast, isogenic HAP1 cell lines mutated at the *MLH1*, *PMS2*, *MSH2*, *EXO1* and *MSH3* loci displayed higher PE levels, ranging from 2 to 6.8-fold (**Figure 1A**). Disruption of other DNA repair pathways had little or no impact on PE efficiency. This finding clearly indicates that MMR

108 functions to inhibit PE. Of all MMR genes targeted in the screen, only the loss of MSH6 failed
109 to increase PE efficiency.

110 The MMR pathway evolved to correct base/base mispairs and small indels arising in DNA
111 during replication and recombination. To initiate repair, these lesions are recognised by the
112 heterodimers MutSα (MSH2-MSH6) or MutSβ (MSH2-MSH3). Whereas MutSα recognises
113 base/base mismatches and indels of 1-2 nucleotides, larger indels are recognised by MutSβ
114 (Drummond et al., 1995; Palombo et al., 1995, 1996; Acharya et al., 1996; Gradia et al., 1997).
115 Substrate binding brings about an ATP-dependent conformational change of the MutS
116 complexes and recruitment of the MutLα (MLH1-PMS2) (Li and Modrich, 1995) or MutL
117 β (MLH1-MLH3) (Lipkin et al., 2000) heterodimers. Assembly of the MutL complex together
118 with RFC and PCNA (Pluciennik et al., 2010), bound at a pre-existing strand discontinuity
119 (either a nick or a free 3' terminus), activates cryptic endonucleases of the PMS2 or MLH3
120 proteins, which then introduce additional DNA single-strand breaks (SSBs) into the
121 discontinuous DNA strand, in the vicinity of the mismatch. These SSBs act as entry points for
122 EXO1, which degrades the discontinuous strand in a 5' to 3' direction up to, and some distance
123 past, the misincorporated nucleotide(s) (Kadyrov et al., 2006). The resulting gap is filled-in by
124 DNA polymerase δ and the remaining nick is ligated by DNA ligase I (Stojic et al., 2004; Iyer
125 et al., 2006; Fishel, 2015). Since the edit introduced in our screen is a 5 bp deletion, this makes
126 it a substrate of MutSβ, but not MutSα (Palombo et al., 1996), which explains the lack of an
127 effect on editing upon the loss of MSH6 (**Figure 1A**). This result highlights the highly
128 specialised nature of the DNA damage response that functions on different substrates.

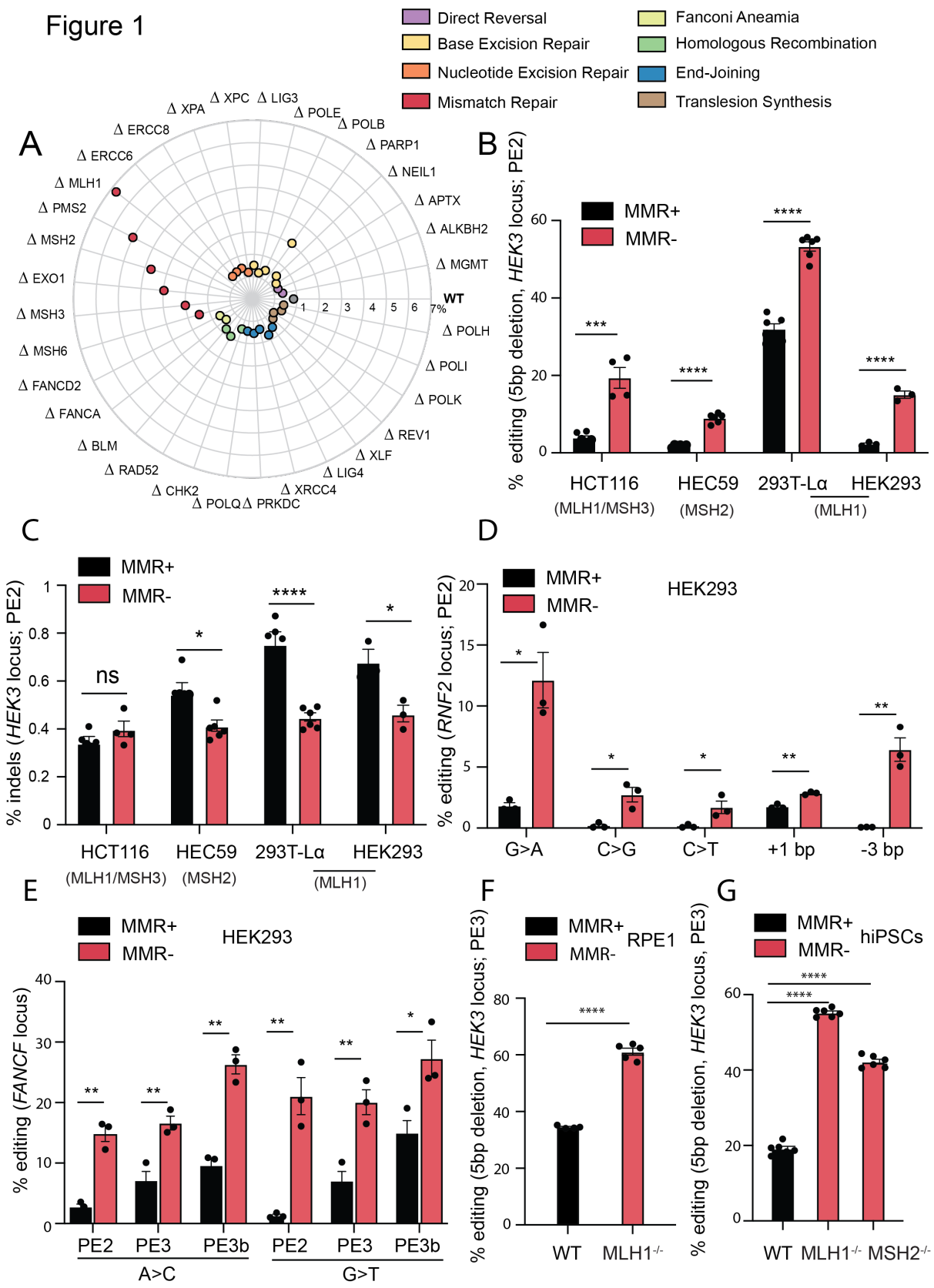


Figure 1: Mismatch repair inhibits prime editing in human cells. A) Genetic screen in 32 HAP1 isogenic knockout cell lines covering different DNA damage repair pathways, as well as their wild-type (WT) counterpart, showing the efficiency of installation of a 5 base pair (bp)

133 deletion in the *HEK3* locus, using PE2. Values correspond to editing efficiency, measured by
134 amplicon sequencing analysis in two independent biological replicates, with two technical
135 replicates each. Each radial line represents an increment of 1%. **B)** PE2 of a 5 bp deletion in
136 the *HEK3* locus in the indicated mismatch repair-deficient cell lines (MMR-), and their
137 respective complemented counterparts (MMR+). For each cell line, the mutated MMR genes
138 are represented. Values correspond to editing efficiency measured by amplicon sequencing
139 analysis in at least three independent biological replicates. **C)** Percentage of indels after
140 inducing a 5 bp deletion in the *HEK3* locus using PE2 in varying mismatch repair-deficient,
141 and their respective complemented, cell lines, in at least three independent biological
142 replicates. Percentage of indels = number of reads containing indels that do not correspond
143 to the desired edit/total number of aligned reads. For each cell line, the correspondent mutated
144 MMR gene is indicated. **D)** PE2 of the indicated types of edits (*RNF2* locus) in HEK293 cells
145 wild-type (MMR+), or knockout for MLH1 (MMR-). Values correspond to editing efficiency,
146 measured by amplicon sequencing analysis in three independent biological replicates. **E)**
147 Efficiency of PE2, PE3 and PE3b after inducing A>C or G>T mutations in the *FANCF* locus.
148 HEK293 wild-type (MMR+) and MLH1 knockout cells (MMR-) were used. Values correspond
149 to editing efficiency, measured by amplicon sequencing analysis in three independent
150 biological replicates. **F)** PE3 efficiency of a 5 bp deletion in the *HEK3* locus in RPE1 wild-type
151 (WT) cells and an isogenic knockout MLH1 cell line (RPE1-MLH1$^{-/-}$), determined by Sanger
152 sequencing and TIDE analysis, for at least three independent biological replicates. RPE1 cells
153 express Cas9(H840A)-RT in a constitutive manner (RPE1 PE2-BSD). **G)** PE3 efficiency in
154 wild-type (WT) human induced-pluripotent stem cells (hiPSCs), as well as isogenic knockouts
155 for MLH1 and MSH2 (MLH1$^{-/-}$, MSH2$^{-/-}$), in three independent biological replicates, with two
156 technical replicates each. Statistical analysis using multiple unpaired t tests. Error bars reflect
157 mean and SEM. Ns p-value non-significant; * p-value < 0.05; ** p-value < 0.01; *** p-
158 value<0.001; **** p-value < 0.0001.

159

160 **Mismatch repair hinders PE2 and PE3 across several human cell lines, genomic loci**
161 **and edit types**

162 To further explore the inhibitory role of MMR in PE, we expanded our investigations to a panel
163 of MMR-deficient human cell lines, alongside their complemented counterparts, in which we
164 measured the editing efficiency and fidelity of the *HEK3* locus. We used the colorectal cancer
165 line HCT116, which is mutated in both *MSH3* and *MLH1*, alongside the MMR-proficient
166 HCT116 cell line complemented with chromosomes 5 and 3 that house the wild-type copies
167 of the two genes, respectively (**Supplementary Figure 1A)** (Koi et al., 1994; Haugen et al.,

168 2008). The endometrial adenocarcinoma cell line HEC59, which is mutated at
169 the *MSH2* locus, was used together with its MMR-proficient counterpart complemented with
170 chromosome 2 that carries the wild-type *MSH2* gene (**Supplementary Figure 1B**) (Umar et
171 al., 1997). Additionally, we used a doxycycline-inducible model of MLH1 deficiency in the
172 embryonic kidney cell line HEK293T (293T-Lα) (**Supplementary Figure 1C**) (Cejka et al.,
173 2003). Finally, we generated an isogenic pair of MLH1 wild-type and knockout HEK293 cells
174 (**Supplementary Figure 1D**).
175
176 We controlled for the transfection efficiencies of all the matched MMR-deficient and proficient
177 cell line pairs and showed that these were comparable, as measured by the percentage of
178 cells transfected with a GFP expressing plasmid (**Supplementary Figure 1E**). We then
179 performed PE2 editing by deleting 5 bp within the *HEK3* locus. All cell lines showed
180 significantly increased PE2 editing (ranging from 1.7 to 6.6-fold) when MMR was ablated,
181 compared to their MMR-proficient counterparts (**Figure 1B**). Importantly, even though PE
182 efficiencies were increased by MMR deficiency, this did not come at the expense of higher
183 indel frequencies within the amplicon region (**Figure 1C**). Indeed, we observed that loss of
184 MMR prevented unwanted indels at the *HEK3* locus in the HEC59, 293T-Lα and HEK293 cell
185 lines (**Figure 1C**).

186 To further investigate the substrates of MMR in PE, we measured the editing efficiencies of a
187 transition (G>A), two transversions (C>G and C>T), a 1 bp insertion and a 3 bp deletion, all
188 within a different endogenous locus, the *RNF2* locus. We found that active MMR significantly
189 diminished the efficiency of all these edits, ranging from 1.6 to 14-fold, using HEK293 cells
190 that lack MLH1, a factor that is part of both the MutLα and MutLβ heterodimers, which together
191 repair base/base mismatches, indels of 1-2 nucleotides and larger indels (**Figure 1D**). These
192 findings were also corroborated in the MLH1/MSH3-deficient HCT116 cell line
193 (**Supplementary Figure 1F**). To test the inhibitory role of MMR on different PE strategies, we
194 measured the efficiency of PE2, PE3 and PE3b on the *FANCF* locus, *via* the installation of
195 either an A>C or a G>T substitution in HEK293 and HCT116 cells. Editing efficiency was
196 improved by MMR deficiency for all types of PE (1.8 -16-fold), albeit to a lesser extent for PE3
197 (**Figure 1E, Supplementary Figure 1G**). Overall, these results show that MMR counteracts
198 PE efficiency across different edits and different genomic loci, in various human cell lines.

199 Since both HCT116 and HEC59 are cancer-derived cell lines that display MMR deficiency, it
200 is possible that the higher levels of PE efficiency are due to cellular adaptation. The human
201 retinal pigmental cell line RPE1 is a non-cancer derived cell line, thus we utilised this for
202 corroborating our findings. PE efficiencies are generally very low in RPE1 wild-type cells

203 (**Supplementary Figure 1H**). To overcome this shortcoming, we developed a lentivirus
204 system for stable delivery of the PE3 system, where RPE1 cells constitutively
205 express Cas9(H840A)-RT (denoted RPE1 PE2-BSD). We generated a CRISPR genetic
206 knockout for the MLH1 factor in this cell line (**Supplementary Figure 1I**) and performed PE3,
207 by transducing both the pegRNA and the nicking sgRNA installing a 5 bp deletion within the
208 *HEK3* locus. We observed an editing efficiency of approximately 35% in WT RPE1 that was
209 further increased to 60% in RPE1-MLH1$^{-/-}$ (**Figure 1F**). We additionally extended our findings
210 to human induced-pluripotent stem cells (hiPSCs), engineered to be deficient for either MLH1
211 or MSH2 (Zou et al., 2021) (**Supplementary Figure 1J**). Wild-type hiPSCs demonstrated
212 20% editing efficiency of a 5 bp deletion in the *HEK3* locus, while the MLH1 and MSH2
213 deficient counterparts displayed an increased efficiency of approximately 55% and 40%,
214 respectively (**Figure 1G**). Overall, these results confirm that the MMR pathway specifically
215 plays a role in counteracting PE.

216 **Mismatch repair factors are recruited to sites of prime editing**

217 To confirm that the MMR proteins are directly involved in the processing of PE intermediates,
218 we determined if they are recruited to sites of ongoing editing marked by Cas9(H840A)-RT.
219 Cas9(H840A)-RT was directed to human repetitive telomeric regions, a strategy that has
220 proven efficient for imaging Cas9 (Chen et al., 2013) (**Figure 2A**). Using this experimental
221 approach, we were able to colocalize TRF1 (an essential component of the telomeric shelterin
222 complex) with catalytically inactive Cas9 (dCas9), as previously described (Wang et al., 2008)
223 and also with Cas9(H840A)-RT (**Supplementary Figure 2A-B**). Therefore, this setup allows
224 for the visualisation of genomic loci undergoing PE, in a pegRNA-dependent manner. Next,
225 we used this system in U2OS cells to express Green Fluorescent Protein (GFP), or GFP-
226 tagged MMR proteins, as well as two additional DNA repair proteins that do not function in
227 MMR (DDB2 that functions in nucleotide excision repair and 53BP1 that promotes non-
228 homologous end-joining). We observed that 65% of MLH1-GFP foci and 25% of MSH2-GFP
229 foci colocalised with Cas9(H840A)-RT foci (**Figure 2B-C**). Importantly, we did not observe
230 colocalisation of either DDB2-GFP or 53BP1-GFP foci and Cas9(H840A)-RT foci (**Figure 2B-**
231 **C**). Furthermore, by using an antibody against MLH1 (**Supplementary Figure 2C-D**) we
232 confirmed the localisation of endogenous MLH1 to sites of PE (**Figure 2D-E)**. We found that
233 30% of MLH1 foci colocalised with Cas9, while we did not observe colocalisation when a
234 dCas9, or a sgRNA, were used (**Figure 2D-E**). These findings reveal that intermediates of PE
235 are substrates of MMR and we propose that MMR functions to degrade the invading
236 heterologous strand and thus restore the original DNA sequence.

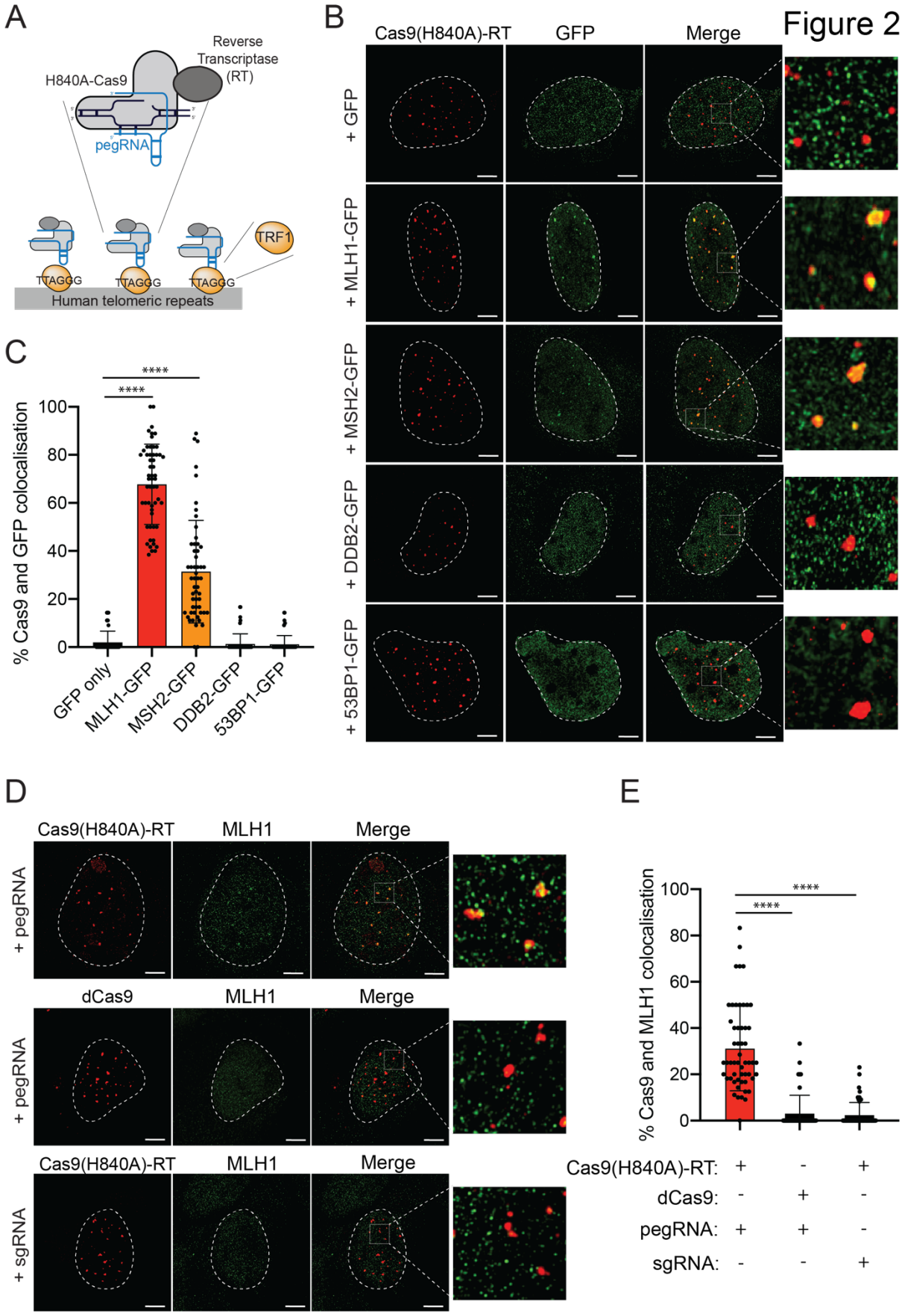


Figure 2
A
Reverse Transcriptase (RT)
H840A-Cas9
pegRNA
TRF1
TTAGGG
TTAGGG
TTAGGG
Human telomeric repeats
B
Cas9(H840A)-RT
GFP
Merge
+ GFP
+ MLH1-GFP
+ MSH2-GFP
+ DDB2-GFP
+ 53BP1-GFP
C
% Cas9 and GFP colocalisation
100
80
60
40
20
0
****
****
GFP only
MLH1-GFP
MSH2-GFP
DDB2-GFP
53BP1-GFP
D
Cas9(H840A)-RT
MLH1
Merge
+ pegRNA
dCas9
MLH1
Merge
+ pegRNA
Cas9(H840A)-RT
MLH1
Merge
+ sgRNA
E
% Cas9 and MLH1 colocalisation
100
80
60
40
20
0
****
****
Cas9(H840A)-RT: + - +
dCas9: - + -
pegRNA: + + -
sgRNA: - - +

237
238

239 **Figure 2: The mismatch repair protein MLH1 localises to sites of active prime editing.**
240 **A)** Scheme of the setup used for imaging. Cas9(H840A)-RT is targeted, through a pegRNA,
241 to human telomeric repetitive regions. TRF1 is a telomeric protein that binds to these regions.
242 **B)** Representative super-resolution images of Cas9(H840A)-RT and the indicated GFP-
243 tagged DNA repair proteins, in U2OS cells 24 hours after reverse transfection with GFP or
244 GFP-tagged MLH1, MSH2, DDB2 or 53BP1, as well as a pegRNA targeting telomeric repeats.
245 Data from three biological replicates and at least 50 cells per condition. **C)** Quantification of B
246 indicating colocalization of Cas9(H840A)-RT foci with GFP foci**. D)** Representative super-
247 resolution images of Cas9(H840A)-RT, or dCas9, and MLH1 with a pegRNA or a sgRNA
248 targeting telomeric repeats. Data from three biological replicates and at least 50 cells per
249 condition. **E)** Quantification of D, indicating colocalization of Cas9(H840A)-RT foci with MLH1
250 foci**.** Statistical analysis using multiple unpaired t tests. Error bars reflect mean and SEM. ****,
251 p-value < 0.0001. All scale bars, 5 µm.

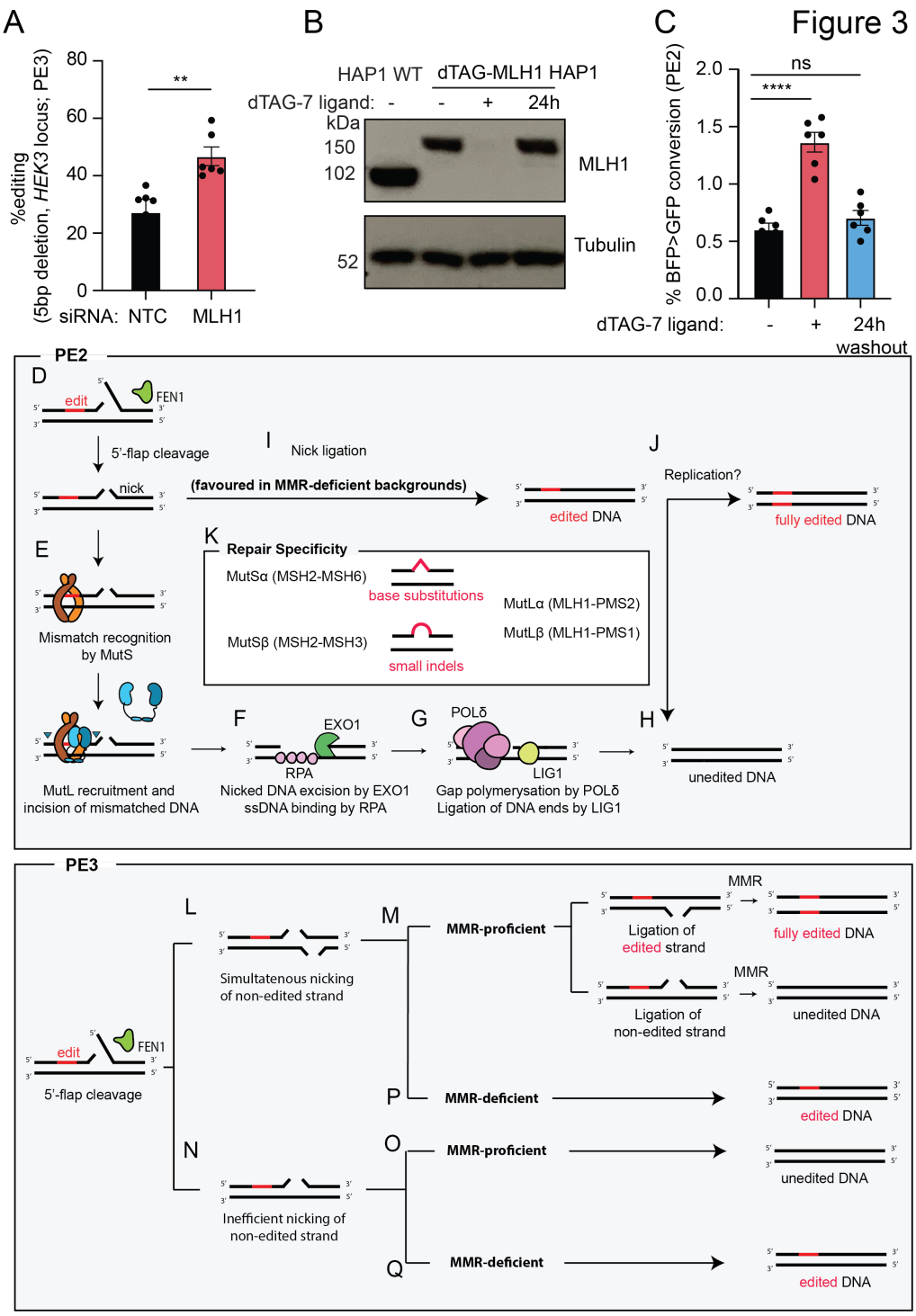
252

253 **Reversible mismatch repair depletion can be exploited to increase prime editing**
254 **efficiency**

255 We next sought to transiently deplete MLH1 as a strategy to improve PE efficiency. Since loss
256 of MMR leads to increased mutational burden and genome instability (Jiricny, 2006), long-
257 term inhibition of MMR is not desirable. Thus, to achieve transient MMR ablation, we depleted
258 MLH1 in HEK293 cells with a pool of siRNAs (**Supplementary Figure 3A**), and subsequently
259 showed that this effectively increased PE efficiency by approximately 2-fold through the
260 generation of a 5 bp deletion in the *HEK3* locus (**Figure 3A**).

261 An alternative approach for achieving transient loss-of-function is through targeted protein
262 degradation. The degradation tag (dTAG) system has proved to be an efficient strategy for
263 rapid and transient ligand-induced targeted protein degradation (Nabet et al., 2018). Using
264 CRISPR-mediated knock-in, we introduced the dTAG into the *MLH1* locus of HAP1 cells,
265 which allowed for the targeted degradation of MLH1 after treatment with the dTAG ligand.
266 Importantly, the protein levels of MLH1 were restored to those found in wild-type cells after 24
267 hours of removal of the ligand (**Figure 3B**). Using a flow cytometry-based readout, in which
268 the pegRNA encodes a 1 bp substitution that converts the Blue-Fluorescent Protein (BFP) to
269 GFP, we observed a 3-fold increase in BFP to GFP conversion upon treatment of the cells
270 with the dTAG ligand and subsequent endogenous degradation of MLH1 (**Figure 3C**). PE
271 efficiency at the *HEK3* locus through a 5 bp deletion, as measured by sequencing genomic
272 DNA, was also significantly increased by 3-fold upon treatment with the dTAG-ligand

273 (**Supplementary Figure 3B**). Importantly, this effect could be rescued to wild-type levels by
274 removing the dTAG-ligand from the medium, thus restoring MLH1 levels (**Figure 3B-C**). Taken
275 together, these results indicate that transient ablation of MMR represents a promising strategy
276 that can be used to increase PE efficiency.



277

278 **Figure 3: Reversible ablation of MLH1 can be exploited to increase prime editing**
279 **efficiency A)** PE efficiency of a 5 bp deletion in the *HEK3* locus in HEK293 cells transfected

280 with non-targeting control (NTC) or MLH1 siRNA pools, two days prior to PE vector delivery.
281 Values correspond to editing efficiency, measured by Sanger sequencing and analysed by
282 TIDE (Brinkman et al., 2014), in three independent biological replicates, with two technical
283 replicates each. **B)** Immunoblot for MLH1 and tubulin in HAP1 cell extracts with ('+') or without
284 ('-') a 1-hour treatment with 500 nM dTAG-7 ligand. Recovery of dTAG-MLH1 expression was
285 measured 24 hours after ligand removal ('24h'). **C)** PE2 efficiency of BFP>GFP conversion in
286 dTAG-MLH1 HAP1 cells, untreated ('-'), or treated with 500 nM of dTAG-7 ligand for 4 days
287 ('+'), or treated with dTAG-7 ligand for 24 hours, followed by its removal for 3 days ('24h
288 washout'). Data measured by flow cytometry for three biological replicates with two technical
289 replicates each. **D-Q)** Schematic model of MMR activity in counteracting PE efficiency. After
290 the cleavage of the non-edited 5'-flap by the flap endonuclease (FEN1), a nick is installed in
291 the edited strand (D). This nick is recognised as a mismatch by the MutS complex, after which
292 MutL is recruited and catalyses incisions that flank the mismatch (E). Exonuclease 1 (EXO1)
293 degrades the incised DNA and Replication protein A (RPA) coats the single-stranded DNA
294 (ssDNA) (F). Polymerase δ fills the gap and Ligase 1 (LIG1) ligates the nick (G). This repair
295 culminates in an unedited DNA molecule (H). If the nick is ligated before mismatch recognition,
296 an heteroduplex DNA is generated, containing the edit in one of the strands (I). The resolution
297 of this heteroduplex potentially relies on replication (J). In PE3, the non-edited strand is
298 simultaneously nicked (L), which directs MMR to repair the mismatch depending on which
299 strand is ligated first (M). However, it is possible that the nicking of the non-edited strand is
300 inefficient, leading to the MMR-mediated removal of the edit (O). In MMR-deficient
301 backgrounds, there is ligation of the heteroduplex without removal of the edit (P,Q). Statistical
302 analysis using unpaired t tests. Error bars reflect mean and SEM. Ns, p-value non-significant;
303 **, p-value < 0.01;****, p-value < 0.0001.

304 Discussion

305 Here we show that the MMR pathway counteracts PE efficiency and fidelity, across different
306 human immortalised and induced pluripotent stem cell lines, genomic loci and edit types.
307 Although the role of MMR in PE had not been addressed experimentally, it was hypothesised
308 to be required for the resolution of the heteroduplex DNA, thus promoting repair of the non-
309 edited strand by utilising the edited strand as template (Petri et al., 2021; Scholefield and
310 Harrison, 2021). Our results provide clear evidence to the contrary, namely that the MMR
311 system functions on the PE intermediate by degrading the invading, RT-synthesised strand
312 to restore the original sequence. This outcome conforms to our understanding of the molecular
313 mechanism of MMR as gleaned from in vitro systems that made use of circular heteroduplex
314 substrates and extracts of human cells (Holmes et al., 1990; Thomas et al., 1991). On these

315 substrates, activation of MMR was strictly dependent on the presence of two factors: a
316 mismatch and a pre-existing nick in one strand that was less than 1 kb distant. The repair
317 process was then directed to the nicked strand. Following the extrapolation of these insights
318 into a cellular setting, MMR, activated by misincorporated nucleotides during replication, would
319 be initiated either by the mismatch/indel and either end of an Okazaki fragment in the lagging
320 strand, or the 3'-end of the primer of the leading strand. During recombination of homologous
321 but non-identical fragments, MMR would be initiated by the heterology (mismatch or indel)
322 between the invading donor and the recipient DNA strands, with the 3'-terminus of the invading
323 strand acting as the signal required to activate the MutL endonucleases.
324
325 We speculate that PE2 resembles the latter mechanism, whereby the RT-synthesised 3' flap
326 would displace the 5' terminus of the Cas9(H840A)-RT-generated nick. This would give rise
327 to a 5' flap, which could be cleaved off by one of several structure-specific endonucleases
328 (SLX1, FAN1, DNA2) and finally trimmed by flap endonuclease 1 (FEN1) (**Figure 3D**). Binding
329 of MutS at the mismatch and its interaction with MutL and RFC/PCNA bound at the 3'-terminus
330 (not shown) would activate the MutL nickase to generate additional incisions flanking the edit
331 (**Figure 3E**). Exonuclease 1 (EXO1) would then degrade the discontinuous strand to generate
332 a long single-stranded gap bound by Replication Protein A (RPA) (**Figure 3F**). Finally,
333 polymerase δ (POL δ) would fill the gap and Ligase 1 (LIG1) ligate the nick (**Figure 3G**). This
334 process would result in the removal of the edit (**Figure 3H**). Ligation of the nick (**Figure 3D**)
335 prior to MMR activation (**Figure 3I**) would generate an heteroduplex with one edited and one
336 non-edited strand (**Figure 3I**), which would be refractory to MMR, and would persist until
337 replication, which would give rise to 50% progeny carrying the edit and 50% non-edited
338 (**Figure 3J**). The path in **Figure 3 D-I-J** would be favoured in the absence of MMR, thus
339 accounting for the increased yield of edited alleles in MMR-deficient backgrounds.
340
341 Importantly, our data confirm the results of biochemical characterisations of the substrate
342 specificities of MutSα and MutSβ (Drummond et al., 1995; Palombo et al., 1995, 1996;
343 Acharya et al., 1996; Gradia et al., 1997), which showed that the former recognises
344 preferentially base/base mismatches and indels of 1-2 nucleotides, whereas the latter binds
345 to larger indels (**Figure 3K**). Based on these findings, deletion of *MSH6* should have failed to
346 affect the outcome of PE using a 5 bp deletion, which was indeed the case, as we report here
347 (**Figure 1A**).
348
349 Besides types of edits, different PE strategies are likely to be impacted by MMR activity to
350 different degrees. PE3 was developed as a more efficient PE strategy, in which both edited
351 and non-edited DNA strands are nicked (**Figure 3L**). When nicks are present in both strands,

352 the nick nearer the mismatch will be preferentially deployed by the MMR system, but the
353 excision will destabilize the duplex and may lead to a DNA DSB, which explains the increased
354 presence of indels in the final outcome of PE3 (Anzalone et al., 2019). If one strand of the
355 heteroduplex is ligated first, MMR is directed to the nicked strand. Therefore, ligation of the
356 edited strand directs MMR to repair the non-edited strand leading to editing on both strands
357 of the DNA heteroduplex, whereas ligation of the non-edited strand results in an unedited DNA
358 molecule (**Figure 3M**). These outcomes rely on the assumption that the nicking sgRNA acts
359 as efficiently as the pegRNA. This might not always be the case as some DNA molecules
360 might have been edited and nicked by the pegRNA only (**Figure 3N**). This would lead to the
361 same outcome as PE2, which is removal of the edit by MMR (**Figure 3O**). In PE3, as well as
362 PE2, cells that lack functional MMR ligate the heteroduplex DNA without removal of the edit
363 (**Figure 3P-Q**). Thus, we propose that MMR activity counteracts PE3 efficiency, as well as
364 PE2, albeit to a lesser extent. This difference is due to the loss of a clear discrimination signal
365 of which strand to repair, created by the nick. In PE3b, the nicking of the non-edited strand is
366 designed to happen only after the integration of the edit. Hence, we propose that the MMR
367 dependency of PE3b is the same as PE2.
368
369 It remains to be seen what the size limitation of PE-generated indels is, that are addressed by
370 MMR. While our work was under revision, two other studies described the suppression of PE
371 efficiency by MMR activity and extensively characterised the types of edits that are efficiently
372 repaired by this pathway. Chen and colleagues showed that MMR involvement decreases with
373 increasing indel size and that G/C to C/G edits, which form C:C mismatches, are less
374 frequently removed by MMR factors (Chen et al., 2021). Koeppel and colleagues
375 systematically measured the insertion efficiency of indels ranging from 1-69 bp in length
376 (Koeppel et al., 2021). The authors observed an overall increase of insertion efficiency upon
377 MMR depletion, with the greatest difference seen for indels 1-4 bp long. These results agree
378 with the known substrate specificities of MutSα and MutSβ (Drummond et al., 1995; Palombo
379 et al., 1995, 1996; Acharya et al., 1996; Gradia et al., 1997). Given that one of the most
380 promising applications of PE includes insertion, deletion or replacement of large sequences
381 of DNA, for example to tag endogenous loci within the genome (Anzalone et al., 2019; Ioannidi
382 et al., 2021), how these lesions are processed and how their insertion efficiency can be
383 augmented, remains of substantial future interest.
384
385 Our findings suggest that the improvement in PE efficiencies in the absence of MMR does not
386 come at the cost of generation of undesirable indels around the edit site (**Figure 1C**). However,
387 MMR deficiency brings about a mutator phenotype, which will severely limit the utility of PE
388 protocols that make use of long-term MMR inactivation. This deleterious outcome might be

389 substantially reduced by interfering with MMR transiently. Our results suggest that targeting
390 MMR factors with siRNA or protein degradation technologies, such as proteolysis targeting
391 chimeras (PROTAC), represent promising approaches to improve PE efficiencies. Another
392 exciting approach would be to interfere with MMR solely at the edit site, similarly to what had
393 been described for improving the efficiency of HDR (Charpentier et al., 2018; Rees et al.,
394 2019). While our article was under revision, two new PE strategies were described, PE4 and
395 PE5, that rely on co-expressing dominant negative MLH1 fragments with the PE2 and PE3
396 machineries, respectively (Chen et al., 2021). The authors also reported that pegRNAs
397 encoding contiguous silent or benign mutations around the intended edit function to evade
398 recognition and repair by MMR (Chen et al., 2021). This strategy has the potential to improve
399 PE efficiency without the increase in mutational burden that is associated with long-term MMR
400 loss.
401
402 Together with recent reports (Chen et al., 2021; Koeppel et al., 2021), our data shed new light
403 on the molecular mechanism of a new and highly promising genome editing technology. We
404 have shown that the MMR pathway inhibits PE efficiency by physically localising to edit sites
405 and promoting their reversion to non-edited sequences. However, the variability in PE
406 observed across cell lines cannot be explained solely by the involvement of MMR and other
407 factors such as cell cycle stage (Wang et al., 2021) or cellular metabolism might also be
408 contributing factors. Hence, further studies are warranted to identify alternative cellular
409 determinants that might limit or promote the use of this technology. The advancement in
410 knowledge reported here can be applied to further the development of prime editors, as well
411 as in the design of novel therapeutic strategies.
412
413 **Methods**
414
415 **Plasmids and oligos**
416 DNA oligos were obtained from Integrated DNA Technologies (IDT) unless otherwise noted.
417 pCMV-PE2 was a gift from David Liu (Addgene plasmid # 132775). pLenti PE2-BSD was a
418 gift from Hyongbum Kim (Addgene plasmid # 161514) (Kim et al., 2021). pU6-pegRNA-GG-
419 acceptor was a gift from David Liu (Addgene plasmid # 132777). PegRNAs were cloned into
420 the pU6-pegRNA-GG-acceptor using BsaI Golden Gate assembly (NEB), following the
421 manufacturer's instructions. sgRNAs utilised in PE3 and PE3b experiments were cloned in the
422 lenti-sgRNA puro vector, using BsmBI Golden Gate assembly (NEB), following the
423 manufacturer's instructions. lenti-sgRNA puro was a gift from Brett Stringer (Addgene
424 plasmid# 104990) (Stringer et al., 2019). lenti-sgRNA neo was a gift from Brett Stringer

425 (Addgene plasmid # 104992) and it was used to clone the sgRNA utilised in PE3 experiments
426 in RPE1 PE2-BSD cells.
427 For immunofluorescence experiments, the dCas9 plasmid was a gift from David Segal
428 (Addgene plasmid # 100091) (O'Geen et al., 2017) and the pSLQ1651-sgTelomere(F+E) was
429 a gift from Bo Huang & Stanley Qi (Addgene plasmid # 51024) (Chen et al., 2013). Additionally,
430 the following plasmids were used: pCMVTet-eGFP-MLH1, pCMVTet-eGFP-MSH2, pEGFP-
431 C1-53BP1. pLenti6.3 WT GFP-DDB2 was also used and it was a gift from Dr. A. Pines (Pines
432 et al., 2012). pmaxGFP™ (Lonza) was used for immunofluorescence experiments, as well as
433 to test transfection efficiency. pCRIS-PITChv2-BSD-dTAG (BRD4), used for the generation of
434 dTAG expressing cells, was a gift from Dr. Georg Winter.
435 BFP-positive cells were generated using the BFP dest clone plasmid. BFP dest clone was a
436 gift from Jacob Corn (Addgene plasmid # 71825) (Richardson et al., 2016).
437 Sequences of sgRNA, pegRNA constructs, as well as primers for genomic DNA amplification
438 are listed in **Supplementary Data 2**. The pegRNA targeting telomeres included a stem loop
439 extension as described in (Chen et al., 2013). All plasmids for mammalian cell experiments
440 were purified using the Plasmid Plus Midi Kit (Qiagen) or the Spin Miniprep Kit (Qiagen), both
441 including endotoxin removal steps.
442 For virus production, the psPAX2 and VSV.G packaging virus were used. psPAX2 was a gift
443 from Didier Trono (Addgene plasmid # 12260). VSV.G was a gift from Tannishtha Reya
444 (Addgene plasmid # 14888).
445
446 **Construction of plentipegRNAPuro vector**
447 The plentipegRNAPuro vector was generated as follows. The lenti-sgRNA puro vector was
448 digested with EcoRI for 2h at 37°C followed by digestion with BsmBI for 2 hours at 55°C and
449 treatment with 4 µl of rSAP (NEB) for 1 hour at 37°C. The mRFP and terminator sequence
450 present in the pU6-pegRNA-GG-acceptor was PCR amplified with a forward primer converting
451 the BsaI cut site to BsmBI and with the reverse primer containing an EcoRI cut site. The PCR
452 product was digested with BsmBI and EcoRI as above. The vector and digest were both
453 purified using gel extraction using the Wizard® SV Gel and PCR Clean-Up System (Promega)
454 and ligated using T4 ligase (NEB) for 1h at room temperature. In order to allow for Golden
455 gate cloning using BsmBI, the BsaI cut site present in the newly assembled vector was
456 converted to a BsmBI cut site using the Q5 Site- Directed Mutagenesis kit (NEB).
457
458 **Lentiviral production and transduction**
459 Lentiviral production was achieved by plating $5x10^6$ xLenti™ cells (Oxgene) in a 10-cm dish
460 transfected one day post seeding with packaging plasmids (1 µg VSV.G, 2 µg psPAX2 and 4

461 µg of transfer plasmid using PEI (Sigma-Aldrich). Virus containing supernatant was collected
462 72 hours post transfection, cleared by centrifugation and stored at -80ºC.
463 Cell transduction was performed using spin-infection as follows. $0.5x10^6$ cells were mixed in a
464 well of a 12-well plate with varying concentrations of supernatant containing viral particles and
465 8 µg/ml of polybrene (Sigma) which was then centrifuged at 2,000 rpm for 30 minutes at 30°C.
466
467 **Mammalian cell culture**
468 All cells were grown at 3% oxygen at 37ºC and routinely checked for possible mycoplasma
469 contamination. Human HAP1 cells were obtained from Horizon Discovery and were grown in
470 Iscove's Modified Dulbecco's Medium (IMDM) (Gibco), containing L-glutamine and 25 nM
471 HEPES and supplemented with 10% Fetal Bovine Serum (FBS) (Gibco) and 1%
472 Penincillin/Streptomycin (P/S) (Sigma-Aldrich). U2OS and HEK293 cells were purchased from
473 ATCC cell repository and cultured in DMEM (Gibco), supplemented with 10% FBS and 1%
474 P/S. HEC59, wild-type and complemented with chromosome 2, were cultured in F12 DMEM
475 with 10% FBS and 1% P/S. HEC59 complemented cells were cultured with 400 µg/mL of
476 geneticin (G418, Gibco). HCT116 cells, wild-type and complemented with both chromosomes
477 3 and 5, were cultured with McCoy's 5A medium (Gibco), with 10% FBS and 1% P/S. HCT116
478 cells complemented with chromosomes 3 and 5 were cultured with 400 µg/mL geneticin
479 (G418, Gibco) and 6 ug/mL blasticidine (Invivogen). 293T-Lα were cultured in DMEM medium
480 (Gibco) supplemented with 10% FBS or Tet-system approved FBS (Takara Bio), 1% P/S, 100
481 µg/mL zeocin (Gibco) and 300 µg/mL hygromycin (Gibco). 293T-Lα were grown in doxycycline
482 (1 µg/mL) for 7 days before any experiment, to completely deplete MLH1 expression.
483 Doxycycline was replenished in the medium every 2 days. RPE1 cells were a gift from the
484 Jackson lab (Gurdon Institute, Cambridge, UK) and cultured in F12 DMEM with 10% FBS and
485 1% P/S. iPSCs (WT, MLH1 and MSH2-deficient) were a gift from the Nik-Zainal lab (University
486 of Cambridge, UK) and cultured on non-tissue culture treated plates (Stem Cell Technologies)
487 pre-coated with 10 µg/mL Vitronectin XF (Stem Cell Technologies) in TeSR-E8 medium (Stem
488 Cell Technologies). The medium was changed daily and the cells were passaged every 4–8
489 days depending on confluency using Gentle Cell Dissociation Reagent (Stem Cell
490 Technologies). 10 µM of ROCKi (Stem Cell Technologies) was added to the medium
491 whenever passaging or thawing iPSCs.
492
493 **Generation of PE2-BSD RPE1 and HEK293T cells**
494 RPE1 and HEK293T cells were transduced at a low multiplicity of infection with the pLenti-
495 PE2-BSD vector and selected two days post transduction with blasticidin (10 µg/ml).
496 Transduced cells were then single cell sorted into 96 well plates and single colonies isolated

497 following 2-3 weeks of clonal expansion. Cas9(H840A)-RT expression was confirmed by
498 immunoblotting.
499
500 **Generation of MLH1 isogenic knockout cell lines**
501 MLH1 knockouts were generated in RPE1 PE2-BSD and HEK293 cell lines by nucleofection
502 of *S.pyogenes* Cas9 together with an in-vitro transcribed sgRNA. Recombinant Cas9
503 containing a nuclear localization sequence and a C-terminal 6-His tag was purchased from
504 Integrated DNA Technologies (#1081059). The sgRNA targeting MLH1 (**Supplementary Data**
505 **2**) was designed utilizing the VBC score tool (https://www.vbc-score.org/). T7 in vitro
506 transcription was performed using HiScribe (NEB E2050S), using PCR-generated DNA as
507 template, as previously described here: dx.doi.org/10.17504/protocols.io.bqjbmuin.
508 The 4D-Nucleofector System X-Unit (Lonza) was used for nucleofection. A mixture of 30 pmol
509 of Cas9 and 60 pmol of in vitro transcribed sgRNA was prepared in a final volume of 5 µL of
510 Cas9 buffer (20 mM HEPES-KOH pH 7.5, 150 mM KCl, 10% glycerol) and incubated for 20
511 minutes, room-temperature. 200,000 HEK293 or RPE1 cells were centrifuged (800 g, 8
512 minutes), washed with PBS and resuspended in 15 µL of SF Cell-Line Solution (V4XC-2032,
513 Lonza) or P3 Primary Cell Solution (V4XP-3032, Lonza), respectively. The Cas9-sgRNA
514 mixture was added to the cells to a final volume of 20 µL and transferred to 16-well
515 Nucleocuvette™ strips (Lonza). Pulse was applied utilizing the CM-130 program for HEK293
516 cells and EA-104 for RPE1 cells. After nucleofection, cells were left to recover for 10 minutes
517 at room temperature, after which they were resuspended in 80 µL of pre-warmed medium,
518 transferred to appropriate dishes and kept in culture.
519 Confirmed of knock-out cell lines was performed by Sanger sequencing, through amplification
520 of genomic DNA with appropriate primers (**Supplementary Data 2**). Tracking of indels by
521 decomposition was performed by the tool TIDE (Brinkman et al., 2014). For RPE1 cells, more
522 than 90% of alleles contained an out-of-frame (+1bp) mutation, which allowed for the use of
523 the pooled population. HEK293 cells showed a lower frequency of out-of-frame indels, hence
524 single cell clones were seeded by limiting dilutions into 96-well plates and a clone containing
525 a +1 bp mutation was selected, 2-3 weeks after clonal expansion, for further studies.
526 Abrogation of MLH1 expression was confirmed in both cell lines by immunoblotting.
527
528 **Focused DNA repair genetic screen**
529 CRISPR-Cas9 knockouts of DNA repair genes were generated in collaboration with Horizon
530 Genomics. Sequences of sgRNAs were designed by Horizon Genomics or with the use of
531 http://chopchop.cbu.uib.no/. sgRNA sequences and frameshift mutations can be found in
532 **Supplementary Data 1**.

533 For the genetic screen, 80,000 cells were seeded in technical duplicates in 12-well plates.
534 Cells were transfected the day after with 636ng of pCMV-PE2 and 159ng of the *HEK3* pegRNA
535 inducing a 5 bp deletion, per well. 1.6 uL Lipofectamine 2000 (ThermoFisher Scientific) were
536 used per well, following the manufacturer's instructions. A separate transfection control was
537 performed using 795 ng of the pmaxGFP^{TM} vector (Lonza). Medium containing transfection
538 reagents was removed 16 hours post-transfection. Transfection efficiency was measured 48
539 hours after transfection, by determining the percentage of GFP positive cells by flow
540 cytometry. Genomic DNA was harvested 96 hours post-transfection, using the QUIAmp DNA
541 Blood Mini kit (Quiagen), following the manufacturer's instructions.
542
543 **Transfection and genomic DNA preparation of mismatch repair-deficient cell lines**
544 HEC59, HCT116 and 293T-Lα and HEK293 cells were seeded in 48-well plates in duplicates
545 (50,000 cells/well). Transfections were performed the next day, using 1 µL Lipofectamine 2000
546 (ThermoFisher Scientific) per well, following the manufacturer's instructions. Cells were
547 transfected with 320 ng of the pCMV-PE2 vector, 80 ng of the respective pegRNA and, for
548 PE3 and PE3b, 33.2 ng of the nicking sgRNA, per well. A transfection control was performed
549 in parallel, by transfecting 400 ng per well of the pmaxGFP^{TM} vector (Lonza).
550 iPSCs were seeded in 48-well plates in duplicates (50,000 cells/well). Transfections were
551 performed the next day, using 1 µL Lipofectamine Stem (ThermoFisher Scientific) per well,
552 following the manufacturer's instructions. Cells were transfected with 320 ng of the pCMV-
553 PE2 vector, 80 ng of the respective pegRNA and, for PE3 and PE3b, 33.2 ng of the nicking
554 sgRNA, per well. A transfection control was performed in parallel, by transfecting 400 ng per
555 well of the pmaxGFP^{TM} vector (Lonza).
556 Genomic DNA was extracted 96 hours after transfection, by removing the medium,
557 resuspending the cells in a lysis solution (100 µL DirectPCR Lysis Reagent (Cell) (Viagen
558 Biotech), 76 µL of water and 4 µL Proteinase K) and incubating 45 minutes at 55°C and 45
559 minutes at 85°C.
560
561 **Prime editing in RPE1 cells**
562 Wild-type and MLH1-knockout RPE1 PE2-BSD cells were transduced at a high multiplicity of
563 infection with the plentipegRNAPuro encoding a 5bp deletion in the *HEK3* locus, together with
564 a nicking sgRNA for PE3 cloned in the lenti-sgRNA neo vector (**Supplementary Data 2**). Spin-
565 infection was performed with 500,000 cells/well in a 12-well plate with 8µg/mL polybrene
566 (2,000 rpm, 90 minutes, 32°C). Cells were selected the day after transduction with blasticidine
567 (10 µg/ml), puromycin (2 µg/ml) and G418 (400 µg/ml). Genomic DNA was extracted as
568 described in the section 'Transfection and genomic DNA preparation of mismatch repair-
569 deficient cell lines', 96 hours post-transduction. Antibiotic selection was maintained throughout

570 the entire duration of the experiment. Prime editing efficiency was measured by Sanger
571 sequencing, after amplification of the genomic DNA with appropriate primers (**Supplementary**
572 **Data 2**). Editing efficiency was calculated by sequence decomposition, using TIDE (Brinkman
573 et al., 2014).
574
575 **High-throughput DNA sequencing of genomic samples**
576 Genomic sites of interest were amplified from genomic DNA samples and sequenced on an
577 Illumina Miseq or NextSeq, depending on the number of pooled samples. Amplification primers
578 containing Illumina forward and reverse primers **(Supplementary Data 2)** were used for a first
579 round of PCR (PCR1) to amplify the genomic region of interest. A mixture of staggered forward
580 primers was used to create complexity. PCR1 reactions were performed in a final volume of
581 25 µL, using 0.5 µM of each forward and reverse primers, 1 µL genomic DNA and 12.5 µL of
582 Phusion U Multiplex PCR 2x Master Mix (ThermoFisher Scientific). PCR1 was carried as
583 following: 98ºC 2 min, 30 cycles [98ºC 10 seconds, 61ºC 20 seconds, 72ºC 30 seconds],
584 followed by a final extension of 72 ºC for 7 minutes. Unique Illumina dual index barcode primer
585 pairs were added to each sample in a second PCR reaction (PCR2). PCR2 was performed in
586 a final volume of 25 µL, using 0.5 µM of each unique forward and reverse Illumina barcoding
587 primer pair,1 µL of unpurified PCR1 reaction and 12.5 µL of of Phusion U Multiplex PCR 2x
588 Master Mix. PCR2 was carried as following: 98ºC 2 minntes, 12 cycles [98ºC 10 seconds,
589 61ºC 20 seconds, 72ºC 30 seconds], followed by a final extension of 72ºC for 7 minutes. PCR
590 products were analysed by electrophoresis in a 1% (w/v) agarose gel and purified using
591 magnetic AMPure XP beads (Beckman Coulter), using a ratio of beads:PCR product of 2:1.
592 DNA concentration was measured by fluorometric quantification (Qubit, ThermoFisher
593 Scientific) and sequenced on an Illumina instrument, according to manufacturer's instructions.
594 Sequencing reads were demultiplexed using MiSeq Reporter (Illumina) and alignment of
595 amplicon sequences to a reference sequence was performed using CRISPResso 2 (Clement
596 et al., 2019). CRISPResso2 was ran in standard mode and prime editing yield was calculated
597 as: number of aligned reads containing the desired edit/total aligned reads. Percentage of
598 indels was calculated as: number of aligned reads containing indels that are not the desired
599 edit/ total number of aligned reads.
600
601 **siRNA transfections**
602 The following siRNAs from Dhamacon (used at a final concentration of 100 nM) were used in
603 this study: MLH1 SMARTpool ON-TARGETplus (L-003906-00-0005) and Non-targeting
604 control SMARTpool ON-TARGETplus (D-001810-10-05). siRNA transfections in HEK293 cells
605 were performed using Dharmafect 1 following manufacturer´s instructions. siRNA delivery was
606 performed 48 hours prior to transfection of prime editing vectors.

607

608 **Generation of dTAG-MLH1 HAP1 cell line**

609 A targeting vector encoding for the BSD-dTAG sequence (amplified from pCRIS-PITChv2-
610 BSD-dTAG (BRD4)) surrounded by two 1kb-long homology arms upstream and downstream
611 of the start codon of MLH1 was generated using Gibson assembly (NEB). An in vitro
612 transcribed sgRNA targeting the region spanning the start codon of MLH1 was generated as
613 previously described (Richardson et al., 2018). Cas9 protein (IDT) together with the targeting
614 vector and in vitro transcribed sgRNA were nucleofected into 200,000 haploid cells in 16-well
615 strips, using a 4D Nucleofector (Lonza) and the program DS-118. Three days after
616 nucleofection, 10 µg/ml blasticidin (Invivogen) were added to the culture medium for one week,
617 after which single and haploid clones were sorted into 96-well plates. Clonal haploid
618 populations were grown and validated for correct homology-directed repair by LR-PCR and
619 immunoblot analysis. dTAG-7 (R&D Systems), at the final concentration of 500 nM, was used
620 to test target degradation in the generated clones and all further targeted protein degradation
621 experiments.

622

623 **Generation of dTAG-MLH1 HAP1 BFP-positive cell line**

624 dTAG-MLH1 HAP1 cells were transduced at a low multiplicity of infection with the BFP dest
625 clone plasmid. Spin-infection was performed with 500,000 cells per well in a 12-well plate
626 (2,000 rpm, 30 minutes, 30°C). Cells were cell sorted by fluorescence (BD FACSMelody), one
627 week after transduction.

628

629 **Prime editing in dTAG-MLH1 HAP1 cell line**

630 25,000 dTAG-MLH1 HAP1 BFP-positive cells were seeded in two technical replicates and
631 three biological replicates in 48-well plates, treated or not with dTAG-7 (R&D Systems) at the
632 final concentration of 500 nM, as indicated. The day after seeding, cells were transfected with
633 200 ng of the pCMV-PE2 vector and 50 ng of a pegRNA cloned into the pU6-pegRNA-GG-
634 acceptor vector, encoding a 1bp substitution in BFP, converting it to GFP (**Supplementary**
635 **Data 2**). dTAG-ligand was replenished in the '+' condition and removed from the '24h'
636 condition. Medium was replaced every 24 hours for the entire course of the experiment (96
637 hours), always replenishing dTAG-7 in the '+' condition. Prime editing efficiency was
638 determined by percentage of GFP positive cells, measured by flow-cytometry.

639 Prime editing efficiency of the *HEK3* locus was measured by seeding 25,000 dTAG-MLH1
640 HAP1 cells in two technical replicates and three biological replicates in 48-well plates, treated
641 ('+') or not ('-') with dTAG-7 (R&D Systems) at the final concentration of 500 nM. The day after
642 seeding, cells were transfected with 200 ng of the pCMV-PE2 vector, 50 ng of the *HEK3*
643 pegRNA and 33.5 ng of the *HEK3* nicking sgRNA for PE3 (**Supplementary Data 2**), using 0.5

644 µL of Lipofectamine 2000 (Thermo Fisher Scientific) and following the manufacturer's
645 instructions. Medium was replaced every 24 hours for the entire course of the experiment,
646 always replenishing dTAG-7 in the '+' condition. Genomic DNA was extracted 96 hours after
647 transfection, by removing the medium, resuspending the cells in a lysis solution (100 µL
648 DirectPCR Lysis Reagent (Cell) (Viagen Biotech), 76 µL of water and 4 µL Proteinase K) and
649 incubating 45 minutes at 55 ºC and 45 minutes at 85 ºC. Prime editing efficiency was
650 determined by Sanger sequencing, after amplifying genomic DNA with appropriate primers
651 (**Supplementary Data 2**) and measured by sequence decomposition using TIDE (Brinkman
652 et al., 2014).
653
654 **Immunoblotting**
655 Cell extracts were prepared in RIPA lysis buffer (NEB) supplemented with protease inhibitors
656 (Sigma) and phosphatase inhibitors (Sigma, NEB). Immunoblots were performed using
657 standard procedures. Protein samples were separated by sodium dodecyl sulfate-
658 polyacrylamide gel electrophoresis (SDS-PAGE) (3–8% gradient gels, Invitrogen) and
659 subsequently transferred onto nitrocellulose membranes. Primary antibodies for MLH1
660 (554073, BD Pharmigen), MSH2 (ab52266, Abcam), MSH3 (ab69619, Abcam), Tubulin (3873,
661 Cell Signaling) and ß-Actin (A5060, Sigma) were used at 1:1,000. Secondary antibodies were
662 used at 1:5,000 (HRP-conjugated goat anti-mouse or anti-rabbit IgG from Jackson
663 Immunochemicals). Immunoblots were imaged using a Curix 60 (AGFA) table-top processor.
664
665 **Immunofluorescence**
666 U2OS cells were reverse transfected using PEI (Sigma-Aldrich). 50,000 cells were seeded
667 per well of µ-Slide 8 well (Ibidi) chambered coverslip plates. Pre-extraction was performed
668 using 0.1% Tween in PBS 24 hours after reverse transfection. Cells were then fixed with 4%
669 para-formaldehyde and fixed cells were processed for immunofluorescence using the
670 following antibodies: anti-Cas9 (Cell Signalling, 14697), anti-TRF1 (Abcam, ab1423), anti-
671 MLH1 (ThermoFisher, A300-015A), anti-MSH2 (Bethyl, A300-452A), anti-GFP (Abcam,
672 ab6556). Primary antibodies were diluted 1:500 and secondary antibodies (Alexa Fluor® 568
673 goat anti-mouse and Alexa Fluor® 488 goat anti-rabbit, LifeTechnologies) were diluted
674 1:2,000.
675
676 **Imaging**
677 16-bit fluorescence images were acquired using an Olympus IXplore spinning disk confocal
678 microscope (equipped with the Yokogawa CSU-W1 with 50 µm pinhole disk and a Hamamatsu
679 ORCA Fusion CMOS camera). A 60X oil immersion objective (NA 1,42) in combination with a
680 3.2X magnification lens (equalling 192X total magnification) was used for super-resolution

681 imaging of fixed cells and z-stacks with a 0.24µm slice interval were acquired. These z-stacks
682 were then processed using the Olympus 3D deconvolution software (constrained iterative
683 deconvolution, using automatic background removal and noise reduction, filter using
684 advanced maximum likelihood algorithm and 5 iterations). Finally, "maximum-z" projection
685 images of the deconvoluted z-stacks were generated. For further data analysis, the ImageJ
686 (NIH) distribution FIJI was used. Nuclear foci were counted manually and at least 50 cells per
687 condition were imaged in each experiment. Quantification of the foci was performed manually
688 based on maximum intensity projections.
689

690 **References**

691 Acharya, S., Wilson, T., Gradia, T., Kane, M., Guerrette, S., Marsischky, G., et al. (1996).
692 hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. *Proc. Natl.*
693 *Acad. Sci. U. S. A.* 93, 13629–13634. doi:10.1073/PNAS.93.24.13629.
694 Anzalone, A. V, Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., et al.
695 (2019). *Search-and-replace genome editing without double-strand breaks or donor DNA.*
696 doi:10.1038/s41586-019-1711-4.
697 Bothmer, A., Phadke, T., Barrera, L. A., Margulies, C. M., Lee, C. S., Buquicchio, F., et al.
698 (2017). Characterization of the interplay between DNA repair and CRISPR/Cas9-induced
699 DNA lesions at an endogenous locus. *Nat. Commun.* 8, 13905.
700 doi:10.1038/ncomms13905.
701 Brinkman, E. K., Chen, T., Amendola, M., and Van Steensel, B. (2014). Easy quantitative
702 assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* 42,
703 e168–e168. doi:10.1093/nar/gku936.
704 Cejka, P., Stojic, L., Mojas, N., Russell, A. M., Heinimann, K., Cannavó, E., et al. (2003).
705 Methylation-induced G2/M arrest requires a full complement of the mismatch repair
706 protein hMLH1. *EMBO J.* 22, 2245. doi:10.1093/EMBOJ/CDG216.
707 Charpentier, M., Khedher, A. H. Y., Menoret, S., Brion, A., Lamribet, K., Dardillac, E., et al.
708 (2018). CtIP fusion to Cas9 enhances transgene integration by homology-dependent
709 repair. *Nat. Commun.* 9, 1–11. doi:10.1038/s41467-018-03475-7.
710 Chen, B., Gilbert, L., Cimini, B., Schnitzbauer, J., Zhang, W., Li, G., et al. (2013). Dynamic
711 imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*
712 155, 1479–1491. doi:10.1016/J.CELL.2013.12.001.
713 Chen, P. J., Hussmann, J. A., Yan, J., Knipping, F., Ravisankar, P., Chen, P.-F., et al. (2021).
714 Enhanced prime editing systems by manipulating cellular determinants of editing
715 outcomes. *Cell* 184, 5635-5652.e29. doi:10.1016/J.CELL.2021.09.018.
716 Clement, K., Rees, H., Canver, M. C., Gehrke, J. M., Farouni, R., Hsu, J. Y., et al. (2019).

717 CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat.*
718 *Biotechnol. 2019 373* 37, 224–226. doi:10.1038/s41587-019-0032-3.
719 Drummond, J., Li, G., Longley, M., and Modrich, P. (1995). Isolation of an hMSH2-p160
720 heterodimer that restores DNA mismatch repair to tumor cells. *Science* 268, 1909–1912.
721 doi:10.1126/SCIENCE.7604264.
722 Ferreira da Silva, J., Meyenberg, M., and Loizou, J. I. (2021). Tissue specificity of DNA repair:
723 the CRISPR compass. *Trends Genet.* 0. doi:10.1016/J.TIG.2021.07.010.
724 Fishel, R. (2015). Mismatch repair. *J. Biol. Chem.* 290, 26395–26403.
725 doi:10.1074/jbc.R115.660142.
726 Gaudelli, N. M., Komor, A. C., Rees, H. A., Packer, M. S., Badran, A. H., Bryson, D. I., et al.
727 (2017). Programmable base editing of A•T to G•C in genomic DNA without DNA
728 cleavage. *Nature*. doi:10.1038/nature24644.
729 Gradia, S., Acharya, S., and Fishel, R. (1997). The Human Mismatch Recognition Complex
730 hMSH2-hMSH6 Functions as a Novel Molecular Switch. *Cell* 91, 995–1005.
731 doi:10.1016/S0092-8674(00)80490-0.
732 Gu, S., Bodai, Z., Cowan, Q. T., and Komor, A. C. (2021). Base editors: Expanding the types
733 of DNA damage products harnessed for genome editing. *Gene Genome Ed.* 1, 100005.
734 doi:10.1016/J.GGEDIT.2021.100005.
735 Haugen, A. C., Goel, A., Yamada, K., Marra, G., Nguyen, T. P., Nagasaka, T., et al. (2008).
736 Genetic instability caused by loss of MutS homologue 3 in human colorectal cancer.
737 *Cancer Res.* 68, 8465–8472. doi:10.1158/0008-5472.CAN-08-0002.
738 Holmes, J., Clark, S., and Modrich, P. (1990). Strand-specific mismatch correction in nuclear
739 extracts of human and Drosophila melanogaster cell lines. *Proc. Natl. Acad. Sci. U. S. A.*
740 87, 5837–5841. doi:10.1073/PNAS.87.15.5837.
741 Hussmann, J. A., Ling, J., Ravisankar, P., Yan, J., Cirincione, A., Xu, A., et al. (2021). Mapping
742 the genetic landscape of DNA double-strand break repair. *Cell* 184, 5653-5669.e25.
743 doi:10.1016/J.CELL.2021.10.002.
744 Ioannidi, E. I., Yarnall, M. T. N., Schmitt-Ulms, C., Krajeski, R. N., Lim, J., Villiger, L., et al.
745 (2021). Drag-and-drop genome insertion without DNA cleavage with CRISPR-directed
746 integrases. *bioRxiv*, 2021.11.01.466786. doi:10.1101/2021.11.01.466786.
747 Iyer, R. R., Pluciennik, A., Burdett, V., and Modrich, P. L. (2006). DNA mismatch repair:
748 Functions and mechanisms. *Chem. Rev.* 106, 302–323. doi:10.1021/cr0404794.
749 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A
750 programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.
751 *Science (80-.).* 337, 816–821. doi:10.1126/science.1225829.
752 Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat. Rev. Mol. Cell Biol. 2006 75*
753 7, 335–346. doi:10.1038/nrm1907.

754 Kadyrov, F., Dzantiev, L., Constantin, N., and Modrich, P. (2006). Endonucleolytic function of
755 MutLalpha in human mismatch repair. *Cell* 126, 297–308.
756 doi:10.1016/J.CELL.2006.05.039.
757 Kim, H. K., Yu, G., Park, J., Min, S., Lee, S., Yoon, S., et al. (2021). Predicting the efficiency
758 of prime editing guide RNAs in human cells. *Nat. Biotechnol.* 39, 198–206.
759 doi:10.1038/S41587-020-0677-Y.
760 Koeppel, J., Madli Peets, E., Weller, J., Pallaseni, A., and Liberante, F. (2021). Predicting
761 efficiency of writing short sequences 1 into the genome using prime editing 2 3. *bioRxiv*.
762 doi:10.1101/2021.11.10.468024.
763 Koi, M., Umar, A., Chauhan, D. P., Cherian, S. P., Carethers, J. M., Kunkel, T. A., et al. (1994).
764 Human Chromosome 3 Corrects Mismatch Repair Deficiency and Microsatellite
765 Instability and Reduces N-Methyl-N'-nitro-N-nitrosoguanidine Tolerance in Colon Tumor
766 Cells with Homozygous hMLHl Mutation. *Cancer Res.* 54, 4308–4312.
767 Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A., and Liu, D. R. (2016). Programmable
768 editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*
769 533, 420–424. doi:10.1038/nature17946.
770 Li, G. M., and Modrich, P. (1995). Restoration of mismatch repair to nuclear extracts of H6
771 colorectal tumor cells by a heterodimer of human MutL homologs. *Proc. Natl. Acad. Sci.*
772 *U. S. A.* 92, 1950–1954. doi:10.1073/pnas.92.6.1950.
773 Lin, Q., Zong, Y., Xue, C., Wang, S., Jin, S., Zhu, Z., et al. (2020). Prime genome editing in
774 rice and wheat. *Nat. Biotechnol.* 38, 582–585. doi:10.1038/s41587-020-0455-x.
775 Lipkin, S. M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A. D., Lynch, H. T., et al.
776 (2000). MLH3: A DNA mismatch repair gene associated with mammalian microsatellite
777 instability. *Nat. Genet.* 24, 27–35. doi:10.1038/71643.
778 Liu, Y., Li, X., He, S., Huang, S., Li, C., Chen, Y., et al. (2020). Efficient generation of mouse
779 models with the prime editing system. *Cell Discov.* 6. doi:10.1038/s41421-020-0165-z.
780 Nabet, B., Roberts, J. M., Buckley, D. L., Paulk, J., Dastjerdi, S., Yang, A., et al. (2018). The
781 dTAG system for immediate and target-specific protein degradation. *Nat. Chem. Biol.*
782 *2018 145* 14, 431–441. doi:10.1038/s41589-018-0021-8.
783 O'Geen, H., Ren, C., Nicolet, C. M., Perez, A. A., Halmai, J., Le, V. M., et al. (2017). dCas9-
784 based epigenome editing suggests acquisition of histone methylation is not sufficient for
785 target gene repression. *Nucleic Acids Res.* 45, 9901–9916. doi:10.1093/NAR/GKX578.
786 Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., et al. (1995).
787 GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells.
788 *Science (80-.).* 268, 1912–1914. doi:10.1126/SCIENCE.7604265.
789 Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996).
790 hMutSβ, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA.

Curr. Biol. 6, 1181–1184. doi:10.1016/S0960-9822(02)70685-4.

Petri, K., Zhang, W., Ma, J., Schmidts, A., Lee, H., Horng, J. E., et al. (2021). CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol. 2021*, 1–5. doi:10.1038/s41587-021-00901-y.

Pines, A., Vrouwe, M. G., Marteijn, J. A., Typas, D., Luijsterburg, M. S., Cansoy, M., et al. (2012). PARP1 promotes nucleotide excision repair through DDB2 stabilization and recruitment of ALC1. *J. Cell Biol.* 199, 235–249. doi:10.1083/JCB.201112132.

Pluciennik, A., Dzantiev, L., Iyer, R. R., Constantin, N., Kadyrov, F. A., and Modrich, P. (2010). PCNA function in the activation and strand direction of MutLα endonuclease in mismatch repair. *Proc. Natl. Acad. Sci. U. S. A.* 107, 16066–16071. doi:10.1073/pnas.1010662107.

Rees, H. A., Yeh, W. H., and Liu, D. R. (2019). Development of hRad51–Cas9 nickase fusions that mediate HDR without double-stranded breaks. *Nat. Commun.* 10, 1–12. doi:10.1038/s41467-019-09983-4.

Richardson, C. D., Kazane, K. R., Feng, S. J., Zelin, E., Bray, N. L., Schäfer, A. J., et al. (2018). CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* doi:10.1038/s41588-018-0174-0.

Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L., and Corn, J. E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* 34, 339–344. doi:10.1038/NBT.3481.

Scholefield, J., and Harrison, P. T. (2021). Prime editing – an update on the field. *Gene Ther. 2021 287* 28, 396–401. doi:10.1038/s41434-021-00263-9.

Stojic, L., Brun, R., and Jiricny, J. (2004). Mismatch repair and DNA damage signalling. *DNA Repair (Amst).* 3, 1091–1101. doi:10.1016/j.dnarep.2004.06.006.

Stringer, B. W., Day, B. W., D'Souza, R. C. J., Jamieson, P. R., Ensbey, K. S., Bruce, Z. C., et al. (2019). A reference collection of patient-derived cell line and xenograft models of proneural, classical and mesenchymal glioblastoma. *Sci. Rep.* 9. doi:10.1038/S41598-019-41277-Z.

Sürün, D., Schneider, A., Mircetic, J., Neumann, K., Lansing, F., Paszkowskirogacz, M., et al. (2020). Efficient generation and correction of mutations in human iPS cells utilizing mRNAs of CRISPR base editors and prime editors. *Genes (Basel).* 11. doi:10.3390/genes11050511.

Thomas, D. C., Roberts, J. D., and Kunkel, T. A. (1991). Heteroduplex repair in extracts of human HeLa cells. *J. Biol. Chem.* 266, 3744–3751. doi:10.1016/s0021-9258(19)67858-0.

Umar, A., Koi, M., Risinger, J. I., Glaab, W. E., Tindall, K. R., Kolodner, R. D., et al. (1997). Correction of hypermutability, N-Methyl-N'-nitro-N-nitrosoguanidine resistance, and defective dna mismatch repair by introducing chromosome 2 into human tumor cells with

mutations in MSH2 and MSH6. *Cancer Res.* 57, 3949–3955.

Wang, Q., Liu, J., Janssen, J. M., Tasca, F., Mei, H., and Gonç Alves, M. A. F. V (2021). Broadening the reach and investigating the potential of prime editors through fully viral gene-deleted adenoviral vector delivery. *Nucleic Acids Res.* 49, 11986–12001. doi:10.1093/NAR/GKAB938.

Wang, X., Kam, Z., Carlton, P. M., Xu, L., Sedat, J. W., and Blackburn, E. H. (2008). Rapid telomere motions in live human cells analyzed by highly time-resolved microscopy. *Epigenetics Chromatin 2008 11* 1, 1–19. doi:10.1186/1756-8935-1-4.

Yeh, C. D., Richardson, C. D., and Corn, J. E. (2019). Advances in genome editing through control of DNA repair pathways. *Nat. Cell Biol.* 21(12), 1468–1478. doi:10.1038/s41556-019-0425-z.

Zou, X., Koh, G. C. C., Nanda, A. S., Degasperi, A., Urgo, K., Roumeliotis, T. I., et al. (2021). A systematic CRISPR screen defines mutational mechanisms underpinning signatures caused by replication errors and endogenous DNA damage. *Nat. Cancer* 2, 643–657. doi:10.1038/s43018-021-00200-0.

Data availability

All sequencing data have been deposited in the European Nucleotide Archive (EMBL-EBI; ENA) with the study accession number PRJEB47501.

Competing Interests

The authors declare no commercial or financial relationships that could be construed as potential conflict of interest.

Author Contributions

JFdaS, GO, JJ and JIL conceptualized the study. JFdaS and JIL obtained funding. JFdaS, GO, EA, CK and AM carried out investigations. GT contributed to the confocal microscopy. JFdaS performed analysis and visualization. JIL supervised the study. JFsaS with input from JIL wrote the original draft and all authors reviewed and edited the final manuscript.

Funding

JFdaS is supported by a DOC fellowship from the Austrian Academy of Sciences (ÖAW25035). A.M. is funded by the Austrian Science Fund, grant number P 33024 awarded to J.I.L. The Loizou lab is funded by an ERC Synergy Grant (DDREAMM Grant agreement ID: 855741). This work was funded, in part, by a donation from Mr Benjamin Landesmann. CeMM is funded by the Austrian Academy of Sciences.

Acknowledgments

We would like to thank the VBCF (Vienna, Austria) for all next generation sequencing. We would like to acknowledge Dr Jacob Corn (ETHZ, Zurich, Switzerland) and Dr Richard Sherwood (Harvard Medical School, Boston, USA) for critically reading the manuscript. We would like to acknowledge Dr Markus Schröder (Corn Lab, ETHZ, Zurich, Switzerland) for bioinformatic analysis. We would like to thank Mr Chris Fell (LBI-RUD/CeMM, Vienna, Austria) and Mr Marc Wiedner (CeMM, Vienna, Austria) for technical support. We are thankful to Prof KJ Patel (Weatherall Institute of Molecular Medicine, Oxford, UK) for providing the FANCD2 deficient HAP1 cell line. Prof Nik-Zainal (University of Cambridge, UK) generously provided the MLH1 and MSH2-deficient human induced pluripotent stem cells. The Jackson lab (The Gurdon Institute, University of Cambridge, UK) kindly provided the RPE cells. The Winter Lab (CeMM, Vienna, Austria) provided valuable tools for protein degradation. We are grateful to members of the Loizou lab for helpful discussions and feedback.