BMB Reports – Manuscript Submission

Manuscript Draft

Manuscript Number: BMB-23-055

Title: Genome-wide In-Locus Epitope Tagging of Arabidopsis Proteins using Prime Editors

Article Type: Article

Keywords: Prime editing; in-locus tagging; Arabidopsis; genome editing; CRISPR

Corresponding Author: Pil Joon Seo

Authors: Pil Joon Seo[1,6,*], Sangsu Bae[3,4,5], Cheljong Hong[1,#], Jun Hee Han[2,#], Gue-Ho Hwang[3]

Institution: [1]Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea,
[2]Department of Chemistry, Hanyang University, Seoul 04673, Republic of Korea,
[3]Medical Research Center of Genomic Medicine Institute and [4]Cancer Research Institute and [5]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 03080, Republic of Korea,
[6]Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, Republic of Korea,

Genome-wide *In-Locus* Epitope Tagging of *Arabidopsis* Proteins using Prime Editors

Cheljong Hong[1,7], Jun Hee Han[2,7], Gue-Ho Hwang[3], Sangsu Bae[3,4,5,*], Pil Joon Seo[1,6,*]

[1]Department of Chemistry, Seoul National University, Seoul 08826, Republic of Korea

[2]Department of Chemistry, Hanyang University, Seoul 04673, Republic of Korea

[3]Medical Research Center of Genomic Medicine Institute, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

[4]Cancer Research Institute, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

[5]Department of Biomedical Sciences, Seoul National University College of Medicine, Seoul 03080, Republic of Korea

[6]Plant Genomics and Breeding Institute, Seoul National University, Seoul 08826, Republic of Korea

[7]These authors equally contributed to this study

*Correspondence: Pil Joon Seo (pjseo1@snu.ac.kr) or Sangsu Bae (sbae7@snu.ac.kr)

20 **Abstract:** Prime editors (PEs), which are CRISPR–Cas9 nickase (H840A)–reverse

21 transcriptase fusion proteins programmed with prime editing guide RNAs (pegRNAs),

22 can not only edit bases but also install transversions, insertions, or deletions without

23 both donor DNA and double-strand breaks at target DNA. As the demands for *in-*

24 *locus* tagging are increasing to reflect gene expression dynamics influenced by

25 endogenous genomic contexts, we demonstrated that PEs can be used to introduce

26 the hemagglutinin (HA) epitope tag to a target gene locus, enabling molecular and

27 biochemical studies using *in-locus* tagged plants. To promote genome-wide *in-locus*

28 tagging, we also implemented a publicly available database that designs pegRNAs

29 for *in-locus* tagging of all *Arabidopsis* genes.

30

31 **Keywords:** Prime editing, *in-locus* tagging, *Arabidopsis*, genome editing, CRISPR

32

33

INTRODUCTION

The search-and-replace method for genome engineering, known as prime editing, was recently developed. A practical version of prime editor (PE), PE2, consists of a SpCas9 nickase (nCas9) containing a H840A mutation and an engineered reverse transcriptase of Moloney murine leukemia virus (M-MLV RT) (1). PE2 can be recruited to a desired target site with the help of a prime editing guide RNA (pegRNA). In addition to a 20-bp protospacer sequence guiding PE2 to the target site, the pegRNA has an extension sequence at the 3' end involving two distinct parts: a primer binding site (PBS) that serves as a template for reverse transcription initiation and a RT template encoding intended edit sequences (1). When the PE-pegRNA complex binds to the target DNA, the nCas9 domain nicks the strand containing the protospacer adjacent motif (PAM), and the liberated 3' end hybridizes to the PBS of the pegRNA, generating the intended edit sequences based on the RT template sequences. The newly generated 3' flap is then annealed with the non-edited strand, and it is expected that the unedited 5' flap will be removed by the innate DNA repair machinery in cells. To maximize prime editing efficiency, PE3 employs an additional nicking single guide RNA (sgRNA) for inducing a second nick at the non-edited strand, which alters the flap equilibrium to include a more desired edit. To date, PEs have been used for base substitution, deletion, and short base insertion in various plant species, including rice, wheat, and tomato (2-10). However, investigation involving insertion of large DNA fragments (>25 bp) such as an epitope tag into plant genomes by PEs has been limited.

There is a growing need for *in-locus* epitope tagging in biological research including plants. Because of the difficulties in generation of antibodies that

3

specifically recognize the target proteins, fusion of epitope tags into open reading frame (ORF) regions of transgenes has been widely performed. However, the random insertion of those transgenes limits full reflection of the gene expression dynamics of the endogenous genes, which may involve long-range chromatin interactions between proximal and distal regulatory elements. In this context, genome-wide *in-locus* protein tagging was suggested in yeast, *Caenorhabditis elegans*, fly, and mammalian cells (11, 12). Following other species, the genome tagging project has been launched in mice (13) and rice (14).

However, although *Arabidopsis thaliana* is a model plant that has been most widely investigated for understanding plant gene and protein dynamics, *in-locus* protein tagging has not yet been widely demonstrated. In particular, given the growing evidence supporting that a number of distal elements are involved in gene expression control as well as a coordinated control of gene clusters (11, 12), *in-locus* protein tagging in *Arabidopsis* is essentially required to reflect realistic *in vivo* functions of genes. Thus, the *in-locus* tagging method will provide invaluable resources to further accelerate plant research.

RESULTS AND DISCUSSION

To facilitate *in-locus* tagging in the *Arabidopsis* genome, we employed a codon-optimized PE3 system (Fig. 1A), in which PEs were expressed under the control of Cauliflower Mosaic Virus (CaMV) 35S promoter. *A. thaliana* U6-26 promoter was used for both pegRNA and nicking sgRNA transcription (Fig. 1B). We also implemented a series of mutations with Cas9 in this plant expression vector to build NG-PAM targetable Cas9 (15). As a proof of concept, the *CIRCADIAN CLOCK*

4

82 *ASSOCIATED 1* (*CCA1*) core clock gene was selected to fuse the HA epitope to the

83 3'-end of the gene ORF using PE3 (Fig. 1C). Two pegRNAs were designed to

84 introduce HA tagging right before the stop codon, which contain 15-nt PBS with either

85 48- or 53-nt RT template (Fig. 1C).

86 We first introduced each construct into *Arabidopsis* plants using an

87 *Agrobacterium*-mediated floral dip method (Fig. 2A). For each pegRNA, two nicking

88 sgRNAs for PE3 application were designed, generating four different constructs (Fig.

89 2B). Leaves of T_1 transgenic plants were excised and used to examine *in-locus*

90 tagging by PEs *in planta*. Immunoblot assay was performed for analysis of

91 approximately 50 T_1 individuals, and all transgenic T_1 plants expressed the CCA1-HA

92 fusion protein, regardless of constructs (Fig. 3A). However, editing frequency in each

93 single individual was likely low (lower than 0.1% detected by deep-seq analysis),

94 because HA tagging to endogenous gene locus was rarely detected by PCR analysis.

95 Next, we collected T_2 bulk seedlings (~15 individuals) and performed immunoblot

96 analysis. According to the result, CCA1-HA fusion protein was also detected in all T_2

97 bulk samples (Fig. 3B). This observation indicates that prime editing is undoubtedly

98 achieved, but editing efficiency of current application is extremely low, consistent with

99 previous reports (6-9, 16-18). Given that editing efficiency of PE is likely low

100 especially in plants compared with mammalian cells, PEs should be improved in the

101 future with the avoidance of plant-specific DNA repair systems to ensure a wide

102 range of genome engineering applications in plants.

103 It is also notable that while the transformed PE system was inherited to the

104 next generation, the parental editing outcomes by PEs (i.e., endogenous *CCA1-HA*

105 tagging) were not inherited. Thus, *in-locus* tagging in T_2 bulk seedlings resulted from

5

106 PE editing at T_2 generation. To explain this, we suspected that the transcriptional

107 activity of 35S promoter may be low in germ cells, which can be improved by

108 replacing the promoters. Alternatively, RT activities might be intrinsically low in plant

109 cells, especially in stem cells and germ cells that transmit genetic information to next

110 generations. Several studies have shown that many plant species can protect

111 themselves against RNA viruses and retrotransposon activation presumably by

112 blocking RT activities (19), supporting our assumption. Hence, alternative strategies

113 are required to obtain homozygous PE-edited plants, such as protoplast regeneration

114 of edited somatic cells (20). In parallel, technical improvements to bypass RT

115 inactivation in germ cells are also essential for enhancing PE applications in plants.

116 　　In spite of the low editing efficiency, we asked whether the *in-locus* tagging

117 indeed reflects endogenous gene expression dynamics. To this end, we arbitrarily

118 chose one T_2 line (#3) and examined CCA1 oscillation patterns in T_2 bulk seedlings

119 grown under neutral day condition (12-h light, 12-h dark). Consistent with protein

120 accumulation detected by a native protein antibody (21), PE-mediated HA fusion to

121 the endogenous *CCA1* gene allows diurnal accumulation of CCA1-HA fusion protein

122 (Fig. 3c). Since PE efficiency is low, the sample pool size should be large enough to

123 minimize sampling bias. We also recommend that current PE-assisted *in*-locus

124 tagging method is more suited for qualitative analysis. Taken together, by means of

125 the current strategy using PE3, *in-locus* tagging can be facilitated at least for various

126 biochemical analyses including immunoblotting, immunostaining, co-

127 immunoprecipitation (Co-IP), immunoprecipitation-mass spectrometry (IP-MS), and

128 chromatin immunoprecipitation (ChIP) in the future.

129 　　In addition, considering the growing demands for *in-locus* tagging and the

130 potential application of PEs to genome-wide *in-locus* tagging, we searched for all

131 possible target sites of PEs *in silico* for 48,276 genes. We found that NGG-targetable

132 PE can cover 95.5% and 91.4% for endogenous tagging at the N-terminus and C-

133 terminus, respectively, of genes, while NG-targetable PE covers more than 99.9% for

134 both the N- and C-terminus of genes (Fig. 4). For user convenience, we open the

135 database (DB) that contains pre-designed pegRNAs for *in-locus* tagging of nearly all

136 *Arabidopsis* genes to the public. The DB is freely available at the site,

137 http://www.rgenome.net/pe-tag-database/download_db.

138   Overall, we demonstrated that epitope tagging with prime editing system at

139 3'-end of the gene ORF is a plausible strategy. We also propose here to tag all

140 proteins encoded by the *Arabidopsis* genome by PE-assisted targeted sequence

141 insertion. The *in-locus* tagging by PE allows *in vivo* analyses on *Arabidopsis* proteins

142 that are particularly under the control of native sequence contexts. It would be

143 invaluable for understanding the dynamics, function, and interaction network of

144 *Arabidopsis* proteins, which provide a fundamental basis for plant research.

145

MATERIALS AND METHODS

Plasmid construction

148 To construct PE plasmid, PE cassette was amplified from pCMV-PE2 (Addgene

149 no.132775) and amplified product was inserted into pBAtC (Addgene no.78097),

150 generating pBAtC-NG-PE2 vector. To build NG-PAM targetable PE vector, we

151 introduced same mutations with pX330-SpCas9-NG (Addgene no.117919) in our

152 Cas9 fragment. For introducing pegRNA cassette, oligos representing the target

153 sequences, sgRNA scaffold and 3' extensions were annealed and cloned into pRG2

154　　vector (Addgene no. 104174) with additional AtU6-26 promoter using BsaI to build

155　　AtU6-26p-pegRNA vector. Restriction enzyme-digested fragment encoding AtU6-26p-

156　　pegRNA cassette was inserted into pBAtC-NG-PE2 vector digested with same

157　　restriction enzyme. To construct nicking sgRNA cassette, oligos representing nicking

158　　sequences were annealed and cloned into AarI-digested PE plasmid. Oligos used for

159　　preparing plasmid was designed using Cas-designer (22) and PE-designer (23).

160

Plant transformation and plant growth conditions

161

162　　The *Arabidopsis* lines generated in this study were in the Col-0 genetic background.

163　　*Agrobacterium tumefaciens* strain GV3101 harboring the PE constructs was used for

164　　*Arabidopsis* transformation. The floral-dip method was performed as previously

165　　described (24). Plants were grown under 16-h light, 8-h dark cycles with cool white

166　　fluorescent light (120 μmol photons m^{-2} s^{-1}) at 23 °C. For detecting circadian

167　　oscillation, 10-day-old seedlings entrained under neutral day conditions (12-h light,

168　　12-h dark) were transferred to continuous light (24-h light) conditions.

169

Western blot assay

170

171　　As for T_1 plants, 3rd rosette leaf of 4-week-old transgenic plants was harvested. For

172　　the T_2 plant analysis, ~15 T_2 seedlings grown for 14 days under long day conditions

173　　on Basta-containing MS medium were harvested for Western blot assay. Harvested

174　　plant materials were ground in liquid nitrogen, and total cellular extracts were

175　　suspended in 2× SDS-PAGE sample loading buffer. The protein samples were then

176　　analyzed by SDS-PAGE (10% polyacrylamide gels) and blotted on to Fluoro Trans

177　　paper (PALL life science, BSP0161). The epitope-tagged proteins were

178 immunologically detected using anti-HA antibody (Millipore, 05-904).

179

186

187 **CONFLICT OF INTEREST**

188 The authors declare no financial or non-financial conflicts of interest.

189

190 **REFERENCES**

191 1. Anzalone AV, Randolph PB, Davis JR et al (2019) Search-and-replace

192   genome editing without double-strand breaks or donor DNA. Nature 576, 149-

193   157

194 2. Butt H, Rao GS, Sedeek K, Aman R, Kamel R and Mahfouz M (2020)

195   Engineering herbicide resistance via prime editing in rice. Plant Biotechnol J

196   18, 2370-2372

197 3. Hua K, Jiang Y, Tao X and Zhu JK (2020) Precision genome engineering in

198   rice using prime editing system. Plant Biotechnol J 18, 2167-2169

199 4. Jiang YY, Chai YP, Lu MH et al (2020) Prime editing efficiently generates

200   W542L and S621I double mutations in two *ALS* genes in maize. Genome

201   Biology 21

9

202 5. Lu YM, Tian YF, Shen RD et al (2021) Precise genome modification in tomato

203   using an improved prime editing system. Plant Biotechnology Journal 19, 415-

204   417

205 6. Li HY, Li JY, Chen JL, Yan L and Xia LQ (2020) Precise Modifications of Both

206   Exogenous and Endogenous Genes in Rice by Prime Editing. Molecular Plant

207   13, 671-674

208 7. Lin QP, Zong Y, Xue CX et al (2020) Prime genome editing in rice and wheat.

209   Nature Biotechnology 38, 582-+

210 8. Tang X, Sretenovic S, Ren QR et al (2020) Plant Prime Editors Enable Precise

211   Gene Editing in Rice Cells. Molecular Plant 13, 667-670

212 9. Xu RF, Li J, Liu XS, Shan TF, Qin RY and Wei PC (2020) Development of

213   Plant Prime-Editing Systems for Precise Genome Editing. Plant

214   Communications 1

215 10. Xue C, Qiu F, Wang Y et al (2023) Tuning plant phenotypes by precise, graded

216   downregulation of gene expression. Nature Biotechnology

217 11. Lackner DH, Carre A, Guzzardo PM et al (2015) A Generic Strategy for

218   CRISPR-Cas9-mediated gene tagging. Nature Communications 6

219 12. Meurer M, Duan YQ, Sass E et al (2018) Genome-wide C-SWAT library for

220   high-throughput yeast genome tagging. Nature Methods 15, 598-+

221 13. Jiang J, Yan M, Li DS and Li JS (2019) Genome tagging project: tag every

222   protein in mice through 'artificial spermatids'. National Science Review 6, 394-

223   396

224 14. Lu YM, Ronald PC, Han B, Li JY and Zhu JK (2020) Rice Protein Tagging

225   Project: A Call for International Collaborations on Genome-wide *In-Locus*

226   Tagging of Rice Proteins. Molecular Plant 13, 1663-1665

227 15. Nishimasu H, Shi X, Ishiguro S et al (2018) Engineered CRISPR-Cas9

228   nuclease with expanded targeting space. Science 361, 1259-1262

229 16. Perroud PF, Guyon-Debast A, Veillet F et al (2022) Prime Editing in the model

230   plant *Physcomitrium patens* and its potential in the tetraploid potato. Plant Sci

231   316, 111162

232 17. Wang L, Kaya HB, Zhang N et al (2021) Spelling Changes and Fluorescent

233   Tagging With Prime Editing Vectors for Plants. Front Genome Ed 3, 617553

234 18. Xu W, Zhang CW, Yang YX et al (2020) Versatile Nucleotides Substitution in

235   Plant Using an Improved Prime Editing System. Molecular Plant 13, 675-678

236 19. Ma XN, Zhang XY, Liu HM and Li ZH (2020) Highly efficient DNA-free plant

237   genome editing using virally delivered CRISPR-Cas9. Nature Plants 6, 773-+

238 20. Jeong YY, Lee HY, Kim SW, Noh YS and Seo PJ (2021) Optimization of

239   protoplast regeneration in the model plant *Arabidopsis thaliana*. Plant Methods

240   17

241 21. Wang ZY and Tobin EM (1998) Constitutive expression of the *CIRCADIAN*

242   *CLOCK ASSOCIATED 1 (CCA1)* gene disrupts circadian rhythms and

243   suppresses its own expression. Cell 93, 1207-1217

244 22. Park J, Bae S and Kim JS (2015) Cas-Designer: a web-based tool for choice

245   of CRISPR-Cas9 target sites. Bioinformatics 31, 4014-4016

246 23. Hwang GH, Jeong YK, Habib O et al (2021) PE-Designer and PE-Analyzer:

247   web-based design and analysis tools for CRISPR prime editing. Nucleic Acids

248   Research 49, W499-W504

249 24. Clough SJ and Bent AF (1998) Floral dip: a simplified method for

250 *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant Journal

251 16, 735-743

252

253

254

255
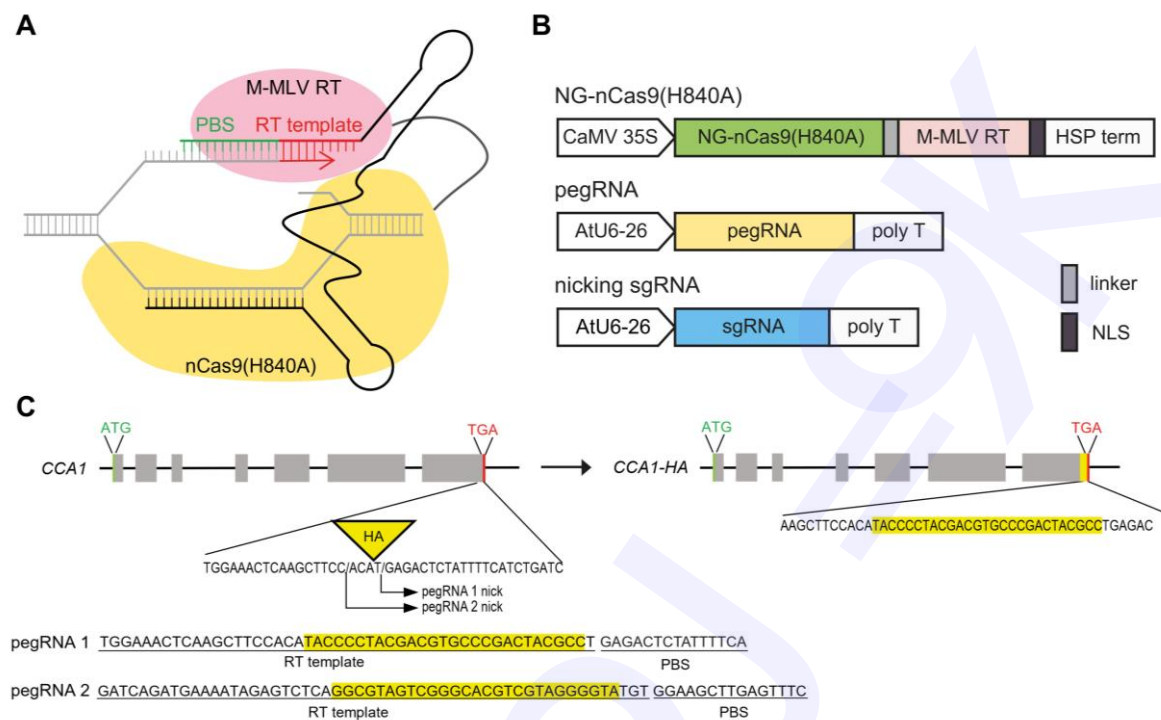
Fig 1. Design of prime editor vector for *in-locus* epitope tagging in *CIRCADIAN CLOCK ASSOCIATED 1* (*CCA1*) gene. (A) Schematic diagram of primer editor application. (B) Vector constructs for endogenous *CCA1-HA* tagging. Prime editor cassette contains NG-PAM targetable nickase Cas9 followed by reverse transcriptase (RT) of Moloney murine leukemia virus (M-MLV) expressed under the control of cauliflower mosaic virus (CaMV) 35S promoter. Both pegRNA and sgRNA are expressed by the AtU6-26 promoter. (C) Schematics of genomic *CCA1* target site and edited *CCA1-HA*. Both pegRNAs contain 15nt-length PBS and 48nt or 53nt-length RT template each. Desired edit outcome preserves HA epitope tag in C-terminus of *CCA1* locus before its stop codon.

266

267

13

268

269

270
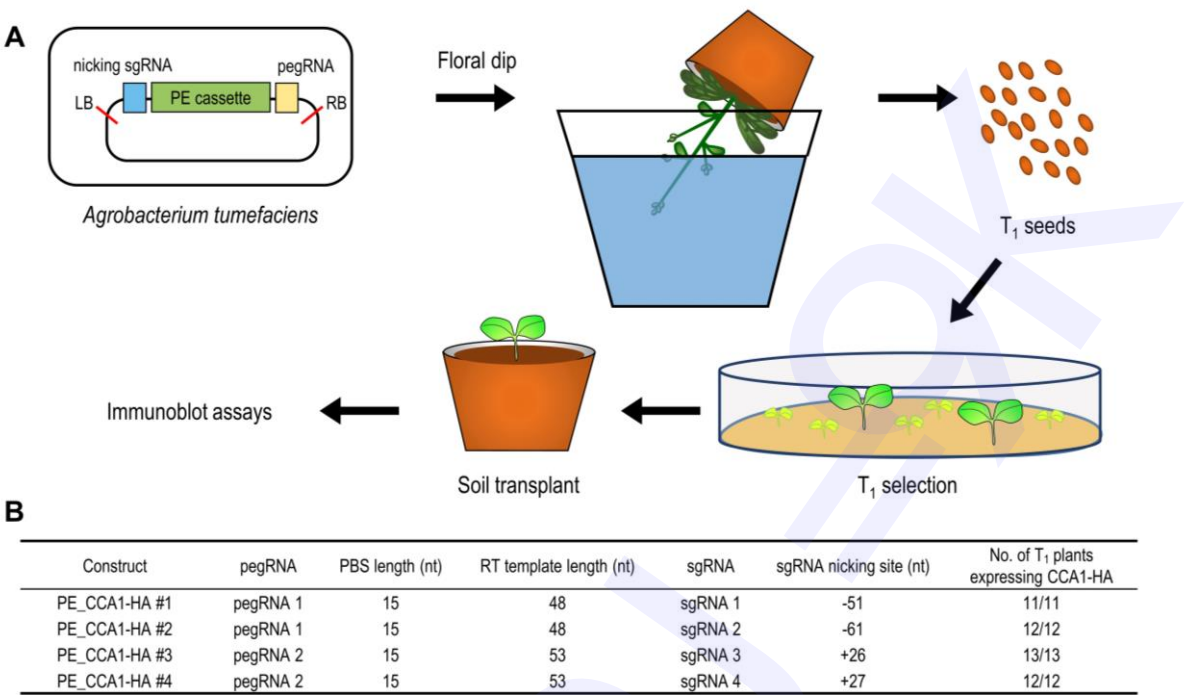
Fig. 2. Schematic design of delivery of prime editor into *Arabidopsis thaliana*. (A) Graphical abstract of *Agrobacterium*-mediated floral dip method. (B) Constructs used for *in-locus* HA tagging. The sgRNA nicking site indicates its distance from the pegRNA nicking site (-, 5'-upstream of pegRNA nicking site; +, 3'-downstream of pegRNA nicking site). Editing frequency was estimated by counting the number of plants expressing the CCA1-HA protein from total plants examined.

14

277
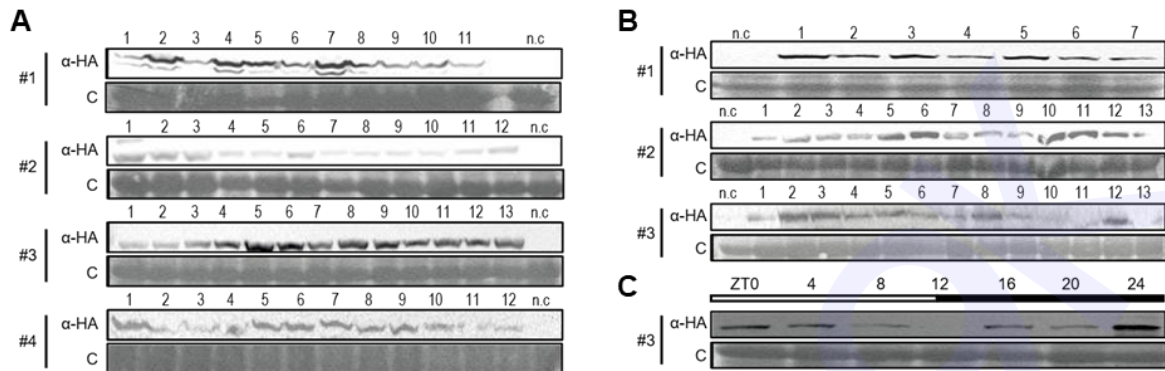
278

279

Fig. 3. Editing outcomes in T_1 individuals and T_2 bulk seedlings. (A) Immunoblot analysis of T_1 individuals. Rosette leaves of 4-week-old plants were excised and used for Western blot analysis. (B) Immunoblot analysis of T_2 bulk seedlings. Fifteen T_2 seedlings were harvested and subjected to Western blot analysis. (C) Diurnal oscillating pattern of CCA1-HA. T_2 seedlings were entrained under neutral day conditions for 2 weeks. Ten seedlings were harvested at each indicated zeitgeber time (ZT) point.

287

288

15

289
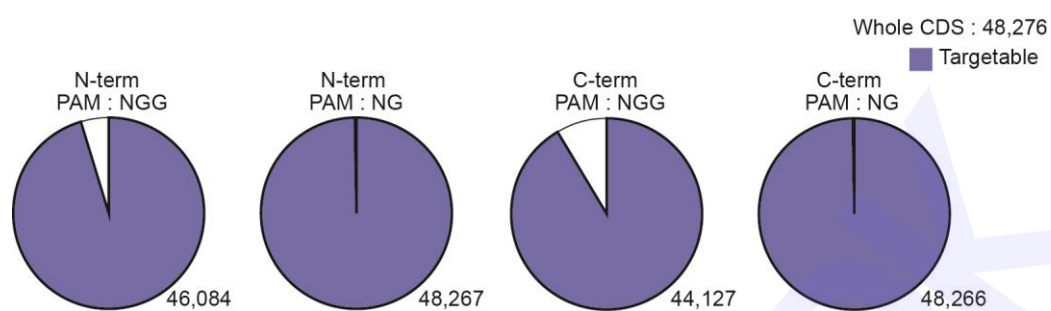
290

291

292 **Fig. 4.** *In silico* analysis of PE targetable genes. The percentages of PE target sites

293 with NGG or NG PAM targetable PEs at the N- and C-terminus of *Arabidopsis* genes.

294