Improving the specificity of adenine base editor using high-fidelity Cas9

Ruisha Hong, Sidi Ma, Feng Wang

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

Corresponding author. *E-mali address*: wangfeng6@mail.sysu.edu.cn (F. Wang).

ABSTRACT

Adenine base editor (ABE) mediates the conversion of A to G in genomic DNA. In human, approximately 47.8% of known pathogenic SNPs can be corrected by A to G conversion, indicating that ABE have tremendous potential in gene therapy. However, the off-target activity of ABE limits its clinical application. ABE off-target activity in DNA is depended on the bonding of *Streptococcus pyogenes* Cas9 (SpCas9) on off-target sites [1, 2]. Therefore, using high-fidelity Cas9 should be able to improve the specificity of ABE. Based on this, we replaced the wild-type SpCas9 in ABE7.10 with four high-fidelity Cas9s to improve its specificity. The analysis of target deep sequencing data demonstrate that the specificity of e-ABE is substantially improved compared to conventional ABE7.10 in four test sites. But the broad editing window of ABE hampers its application, ABE needs to be optimized to get variants with narrow editing window.

KEYWORDS

CRISPR/Cas9, high-fidelity ABE, off-target activity, pathogenic SNPs

1. Introduction

The CRISPR (clustered regularly interspaced short palindromic repeats)-Cas (CRISPR-associated proteins) is an adaptive immune system widely existing in bacteria and archaea, which can cleavage invading foreign nucleic acids directed by short guide RNAs [3, 4]. The system has been engineered as gene editing tools, including a single guide RNA (sgRNA) and Cas9 nucleases [5, 6]. The sgRNA-Cas9 complex cleavages target DNA sequence based on two conditions: (1) compatibility of the protospacer adjacent motif (PAM) with the PAM-interacting domain of Cas9, and (2) complementary of the sgRNA sequence with the target site. However, the nuclease activity of Cas9 may also be triggered when there are mismatches between sgRNA and off-target site [7, 8]. These off-target effects limit its application in genome-editing, especially in clinical.

Several strategies have be used to reduce the off-target activity of Cas9, including reducing the amount of active Cas9 [9], using paired nickases Cas9 (nCas9) [10, 11], fusing Cas9 with specific DNA-binding proteins [12, 13], engineering of the sgRNAs (truncated sgRNAs [14] and hairpin sgRNAs [15]) and so on. In addition, several high-fedility Cas9 variants have been developed, such as enhanced specificity Cas9 (eSpCas9) [16], Cas9-high fidelity (Cas9-HF) [17], evoCas9 [18], HypaCas9 [19] and sniper-Cas9 [20].

Based on CRISPR/Cas9, single base editing system has been developed, including cytosine base editors (CBEs) [21] and adenosine base editors (ABEs) [2]. CBEs enable

C•G to T•A base pair conversion, while ABEs enable A•T to G•C base pair conversion, at a target genomic locus without inducing double strand breaks. ABEs consist of two parts, sgRNA and fusion protein of adenosine deaminase TadA and nCas9. Locating in a target site guided by sgRNA, TadA catalyzes base A deamination to I (inosine) on the non-complementary strand, meanwhile nCas9 nick the complementary strand. Because base I can pair with C, A•T convert to G•C after replication.

In human, about 47.8% of known pathogenic single nucleotide polymorphisms (SNPs) can be corrected by conversion of A•T to G•C (Fig. 3A), indicating ABEs have tremendous potential in gene therapy [2]. However, ABEs can tolerate mismatch between sgRNA and target sequence, inducing off-target effects [22], which limit its application especially in clinical. The using of high-fidelity Cas9 variants in ABE have not been reported, except sniper-Cas9. But sniper-ABE also have off-target activity on some sites [23]. Because the high-fidelity Cas9 variants not always have same off-target sites [18], we need various high-fidelity ABEs. Here, we replaced the wild type SpCas9 in ABE7.10 with four high-fidelity Cas9s (eSpCas9, Cas9-HF, evoCas9 and hypaCas9), named e-ABE7.10, HF-ABE7.10, evo-ABE7.10 and Hypa-ABE7.10, in order to enhance the specificity of ABE.

2. Materials and methods

2.1. Plasmid construction

The conventional ABE7.10 plasmid (pCMV-ABE7.10; Addgene, #102919) was purchased from Addgene, besides we optimized its codon according previous report [24].The sgRNA plasmid pUC57-Cas9-gRNA was synthesized by GENEWIZ. The pCMV-eABE7.10, pCMV-HypaABE7.10 and pCMV-evoABE7.10 plasmids were constructed through site directed mutagenesis by PCR on pCMV-ABE7.10 plasmid backbone, and mutated sites are shown in Fig. 1A. The pCMV-HFABE7.10 plasmid was constructed through Gibson assembly of necessary sequence (synthesized by GENEWIZ) into the pCMV-ABE7.10. Primers used for gRNA inserting into the pUC57-Cas9-gRNA plasmid are shown in Supplementary Table 3. All plasmids were confirmed by Sanger sequencing.

2.2. Human cell culture and transfections

HEK293T cells were cultured and passaged in Dulbecco's modified Eagle's medium (DMEM, Gibco) supplemented with 10% (v/v) fetal bovine serum (FBS, Gibco) and 1% (v/v) penicillin-streptomycin (Gibco). Cell lines were maintained at 37 °C with 5% CO_2. Plasmids for mammalian cell transfections were prepared using an Endo-free Plasmid Mini Kit II (OMEGA). HEK293T (1.5×10^5 cells) were seeded into 24-well Poly-d-Lysine-coated plates (Corning) in the absence of antibiotic. After12-15 hours plating, 750 ng ABE plasmid and 250 ng gRNA plasmid were co-transfected into the cells with Lipofectamine 3000 according to the manufacturer's protocol. Genomic DNA was extracted 5 days after transfection by Tissue DNA Kit (OMEGA).

2.3. DNA amplification and deep sequencing

Primers used for on- and off-target sites PCR amplification were listed in Supplementary Table 3. PCR was performed using Q5 Host Start High- Fidelity 2× Master Mix (NEB), and 150 ng Genomic DNA was chosen as the template. The program of PCR: 98°C for 30 s, then 35 cycles of (98°C for 10 s, 67°C for 10 s, and 72°C for 10 s), followed by a final extension of 2 min at 72°C. PCR products were purified by electrophoresis and deep sequenced using the HiSeq 2x150bp sequencing system (Illumina) as paired-end 150 reads at GENEWIZ.

2.4. Data analysis

Because the length of reads didn't cover the whole fragments sequenced. For data sequenced by NGS platforms, we used FLASH [25] to merge the pair-ends sequence reads to do the following process. The nucleotide constitution analysis were performed with python referencing the python script in a previous paper [2]. Some adjustments were done to fit our experiment requirements.

For editing site analysis, we used the ClinVar data with "fileData = 2019-06-18". We chose data in VCF format with the INFO of "CLNSIG=Pathogenic", i.e. clinical significance for this single variant was pathogenic. Using these data could assess candidates for base editing therapy. The potential window size varied between 1 and 8, and latent editing windows were located in fragment between third base and tenth base. The human genome analyzed is Genome Reference Consortium Human Build 38 (GRCh38). The data and scripts are available from the authors upon reasonable request.

3. Results

We selected the well validated site *HEK4* [2] to test the specificity of the four ABEs. On the three known off-target sites, the specificity of four ABEs increased to 2.5 - 54.5 times, compared with ABE7.10 (Fig. 1 B, C). The on-target editing activity of e-ABE7.10 is comparable to ABE7.10, but three others have varying degrees of reduction compared to ABE7.10. Therefore, we chose e-ABE7.10 for further testing on the other known sites (*HBG2*, *HPRT* and *VEGFA3* [22, 23]). As shown in Fig. 2, the specificity ratios are increased to 1.1 - 44.2 times, except *HPRT* off-target site 2 and 3. This is due to that both e-ABE7.10 and ABE7.10 have no observable base editing at the *HPRT* off-target site2 and 3, but ABE7.10 have a higher base editing activity at *HPRT* on-target than e-ABE7.10. Compared with ABE7.10, the on-target base editing activity of e-ABE7.10 maintained the same level at site *VEGFA3*, but it have some decrease at sites *HBG2* and *HPRT*. Collectively, e-ABE7.10 has a high specificity in a site dependent method.

We find that both e-ABE7.10 and ABE7.10 also have base editing activity at the position 3 and 9 of protospacer, consistent with previous reports [2]. The broad editing windows hamper its application in clinical due to it may contain more than one A in the window. The canonical ABE7.10 original editing window is between third base and tenth base from 5' end to 3'end [2]. Many adenines in this fragment are edited by ABE leading to wrong editing. One solution to this defect is shrinking the editing windows from 8 bases fragments to smaller fragments. But we have no idea where the new editing window will arise in the editing window, every position in the editing window

is possible. As the windows length and location relative to 5' end changes, SNPs' editable situations also alter much.

There are two factors to determine whether one SNP can be corrected using ABE without wrong editing in editing window. Firstly, the longer new editing windows are, the more probably additional non-target adenines will arise in editing windows which is not acceptable for clinical use. Secondly, the longer editing windows are, the more likely the PAMs will find, because target adenine can change its location in editing windows allowing find PAM in other places. For example, for window size of 8, a pathogenic SNP can choose 8 possible distance between target site and PAM sequences. But for window size of 1, the position of target nucleotide is deterministic and the PAM position is exclusive. These two factors are antagonist, so we analysis all the potential target inducing human disease stored in ClinVar datebase [26]. When the editing window is in positions 3-9, 11.7% and 34.8% of pathogenic G to A SNPs can be corrected by spCas9-ABE and xCas9-ABE, respectively (Fig. 3). The xCas9 has a broad PAM range (5′-NG, 5′-GAA, 5′-GAT, and 5′-CAA) [27]. If the editing window is in positions 4-8, the percent will increase to 14.4% and 43.8%. We analyzed the possible ABEs with different editing window 3-10, indicating that the ABE with editing window 4-7 have largest target number in SpCas9-ABE but the window is 4-6 in xCas9-ABE. To summarize, we should engineer ABE variants with different editing window, especially with window 4-6 in xCas9-ABE.

4. Discussion

To reduce or avoid ABE off-target activity, we replaced ABE's original Cas9 with four high-fidelity Cas9. At different loci, e-ABE showed 1.1 to 54.5 times improvement in specificity ratios, compared with ABE7.10. eSpCas9 consists three neutral alanine substitution at positively charged residues within the nontarget strand groove, requiring more stringent Watson-Crick base pairing between sgRNA and the target DNA strand for the single strand formation of nontarget DNA strand [16]. The single DNA strand is needed for base A deamination by ABE [2]. It should also be able to improve the specificity of CBE. Extended sgRNA can reduce ABE off-target activity[23], it may have the same effect in e-ABE7.10, needing further test.

According to recent reports, the off-target sites of ABE is fewer in vitro analysis by whole-genome sequencing [22], and in vivo, the same results were observed in rice [1] and mouse embryos [28]. So we only chose the top three or four off-target sites with relatively strong off-target activity to test the specificity of ABEs. But we should perform more comprehensive test, in order to get more convincing results. At the same time, we realized that the ABE with broad editing window will hamper its application in clinical (Fig. 3), but we do not have an ABE with narrow editing window by now. Engineering ABE variants with narrow window is a more important task. There is still a long way to go and more manpower and material resources should be invested in the field.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgements

References

[1] S. Jin, Y. Zong, Q. Gao, Z. Zhu, Y. Wang, P. Qin, C. Liang, D. Wang, J.L. Qiu, F. Zhang, C. Gao, Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice, Science 364 (2019) 292-295.

[2] N.M. Gaudelli, A.C. Komor, H.A. Rees, M.S. Packer, A.H. Badran, D.I. Bryson, D.R. Liu, Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, Nature 551 (2017) 464–471.

[3] B. Wiedenheft, S.H. Sternberg, J.A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea, Nature 482 (2012) 331-338.

[4] D. Bhaya, M. Davison, R. Barrangou, CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation, Annu. Rev. Genet. 45 (2011) 273-297.

[5] L. Cong, F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, F. Zhang, Multiplex genome engineering using CRISPR/Cas systems, Science 339 (2013) 819-823.

[6] P. Mali, L. Yang, K.M. Esvelt, J. Aach, M. Guell, J.E. DiCarlo, J.E. Norville, G.M. Church, RNA-guided human genome engineering via Cas9, Science 339 (2013) 823-826.

[7] P.D. Hsu, D.A. Scott, J.A. Weinstein, F.A. Ran, S. Konermann, V. Agarwala, Y. Li, E.J. Fine, X. Wu, O. Shalem, T.J. Cradick, L.A. Marraffini, G. Bao, F. Zhang, DNA targeting specificity of RNA-guided Cas9 nucleases, Nat. Biotechnol. 31 (2013) 827-832.

[8] Y. Fu, J.A. Foden, C. Khayter, M.L. Maeder, D. Reyon, J.K. Joung, J.D. Sander, High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells, Nat. Biotechnol. 31 (2013) 822-826.

[9] K.M. Davis, V. Pattanayak, D.B. Thompson, J.A. Zuris, D.R. Liu, Small molecule-triggered Cas9 protein with improved genome-editing specificity, Nat. Chem. Biol. 11 (2015) 316-318.

[10] F.A. Ran, P.D. Hsu, C.Y. Lin, J.S. Gootenberg, S. Konermann, A.E. Trevino, D.A. Scott, A. Inoue, S. Matoba, Y. Zhang, F. Zhang, Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity, Cell 154 (2013) 1380-1389.

[11] P. Mali, J. Aach, P.B. Stranges, K.M. Esvelt, M. Moosburner, S. Kosuri, L. Yang, G.M. Church, CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering, Nat. Biotechnol. 31 (2013) 833-838.

[12] J.P. Guilinger, D.B. Thompson, D.R. Liu, Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification, Nat. Biotechnol. 32 (2014) 577-582.

[13] M.F. Bolukbasi, A. Gupta, S. Oikemus, A.G. Derr, M. Garber, M.H. Brodsky, L.J. Zhu, S.A. Wolfe, DNA-binding-domain fusions enhance the targeting range and

precision of Cas9, Nat. Methods 12 (2015) 1150-1156.

[14] Y. Fu, J.D. Sander, D. Reyon, V.M. Cascio, J.K. Joung, Improving CRISPR-Cas nuclease specificity using truncated guide RNAs, Nat. Biotechnol. 32 (2014) 279-284.

[15] D.D. Kocak, E.A. Josephs, V. Bhandarkar, S.S. Adkar, J.B. Kwon, C.A. Gersbach, Increasing the specificity of CRISPR systems with engineered RNA secondary structures, Nat. Biotechnol. 37 (2019) 657–666.

[16] I.M. Slaymaker, L. Gao, B. Zetsche, D.A. Scott, W.X. Yan, F. Zhang, Rationally engineered Cas9 nucleases with improved specificity, Science 351 (2016) 84-88.

[17] B.P. Kleinstiver, V. Pattanayak, M.S. Prew, S.Q. Tsai, N.T. Nguyen, Z. Zheng, J.K. Joung, High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects, Nature 529 (2016) 490-495.

[18] A. Casini, M. Olivieri, G. Petris, C. Montagna, G. Reginato, G. Maule, F. Lorenzin, D. Prandi, A. Romanel, F. Demichelis, A. Inga, A. Cereseto, A highly specific SpCas9 variant is identified by in vivo screening in yeast, Nat. Biotechnol. 36 (2018) 265.

[19] J.S. Chen, Y.S. Dagdas, B.P. Kleinstiver, M.M. Welch, A.A. Sousa, L.B. Harrington, S.H. Sternberg, J.K. Joung, A. Yildiz, J.A. Doudna, Enhanced proofreading governs CRISPR-Cas9 targeting accuracy, Nature 550 (2017) 407-410.

[20] J.K. Lee, E. Jeong, J. Lee, M. Jung, E. Shin, Y.H. Kim, K. Lee, I. Jung, D. Kim, S. Kim, J.S. Kim, Directed evolution of CRISPR-Cas9 to increase its specificity, Nat. Commun. 9 (2018) 3048.

[21] A.C. Komor, Y.B. Kim, M.S. Packer, J.A. Zuris, D.R. Liu, Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage, Nature 533 (2016) 420-424.

[22] P. Liang, X. Xie, S. Zhi, H. Sun, X. Zhang, Y. Chen, Y. Chen, Y. Xiong, W. Ma, D. Liu, J. Huang, Z. Songyang, Genome-wide profiling of adenine base editor specificity by EndoV-seq, Nat. Commun. 10 (2019) 1-9.

[23] D. Kim, D.E. Kim, G. Lee, S.I. Cho, J.S. Kim, Genome-wide target specificity of CRISPR RNA-guided adenine base editors, Nat. Biotechnol. 37 (2019) 430–435.

[24] M.P. Zafra, E.M. Schatoff, A. Katti, M. Foronda, M. Breinig, A.Y. Schweitzer, A. Simon, T. Han, S. Goswami, E. Montgomery, J. Thibado, E.R. Kastenhuber, F.J. Sanchez-Rivera, J. Shi, C.R. Vakoc, S.W. Lowe, D.F. Tschaharganeh, L.E. Dow, Optimized base editors enable efficient editing in cells, organoids and mice, Nat. Biotechnol. 36 (2018) 888-893.

[25] T. Magoc, S.L. Salzberg, FLASH: fast length adjustment of short reads to improve genome assemblies, Bioinformatics 27 (2011) 2957-2963.

[26] M.J. Landrum, J.M. Lee, M. Benson, G. Brown, C. Chao, S. Chitipiralla, B. Gu, J. Hart, D. Hoffman, J. Hoover, W. Jang, K. Katz, M. Ovetsky, G. Riley, A. Sethi, R. Tully, R. Villamarin-Salomon, W. Rubinstein, D.R. Maglott, ClinVar: public archive of interpretations of clinically relevant variants, Nucleic Acids Res. 44 (2016) D862-868.

[27] J.H. Hu, S.M. Miller, M.H. Geurts, W. Tang, L. Chen, N. Sun, C.M. Zeina, X. Gao, H.A. Rees, Z. Lin, D.R. Liu, Evolved Cas9 variants with broad PAM compatibility and high DNA specificity, Nature 556 (2018) 57-63.

[28] E. Zuo, Y. Sun, W. Wei, T. Yuan, W. Ying, H. Sun, L. Yuan, L.M. Steinmetz, Y. Li, H. Yang, Cytosine base editor generates substantial off-target single-nucleotide variants

in mouse embryos, Science 364 (2019) 289-292.

Figure Legends

Fig. 1. Construction of the high-fidelity ABEs and its base editing analysis. (A) Schematic representation of ABE.10, e-ABE7.10, HF-ABE7.10, evo-ABE7.10 and Hypa-ABE7.10. (B) Base editing efficiencies of ABEs measured by targeted deep sequencing at HEK4 on- and off-target sites in HEK293T cell. Mismatched bases, edited bases and PAM sequences are shown in red, green and blue, respectively. (C) Specificity ratios were shown by the heatmap, calculated by the formula: (high-fidelity ABE7.10 on-target frequency/off-target frequency)/(ABE7.10 on-target frequency/off-target frequency). Means ± SD were from two or three independent experiments.

Fig. 2. Base editing analysis of eABE7.10. Base editing efficiencies of eABE7.10 and ABE7.10 measured by targeted deep sequencing at *HBG2* (A), *HPRT* (B) and *VEGFA3* (C) on- and off-target sites in HEK293T cells. Mismatched bases, edited bases and PAM sequences are shown in red, green and blue, respectively. Dashes indicate presumed RNA bulges. Specificity ratios were shown by the heatmap, calculated by the formula in Fig. 1. Means ± SD were from three independent experiments.

Fig. 3. Pathogenic SNPs can be corrected using ABEs. (A) Base pair changes need to correct pathogenic SNPs in the ClinVar Datebase[26]. The percent of pathogenic SNPs (G to A) can be corrected by SpCas9-ABE (B) and xCas9-ABE (C). The assumed ABEs have different editing window.
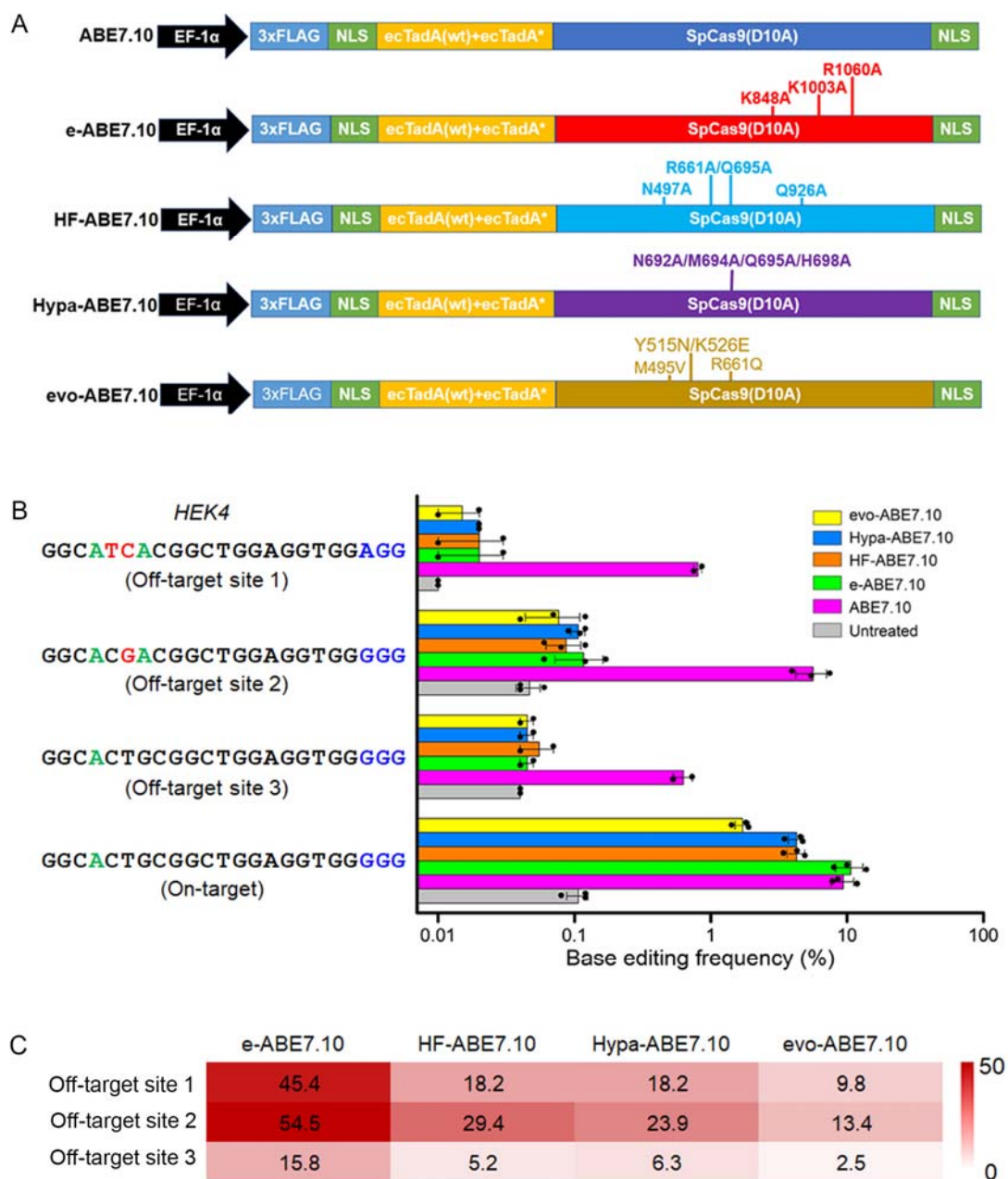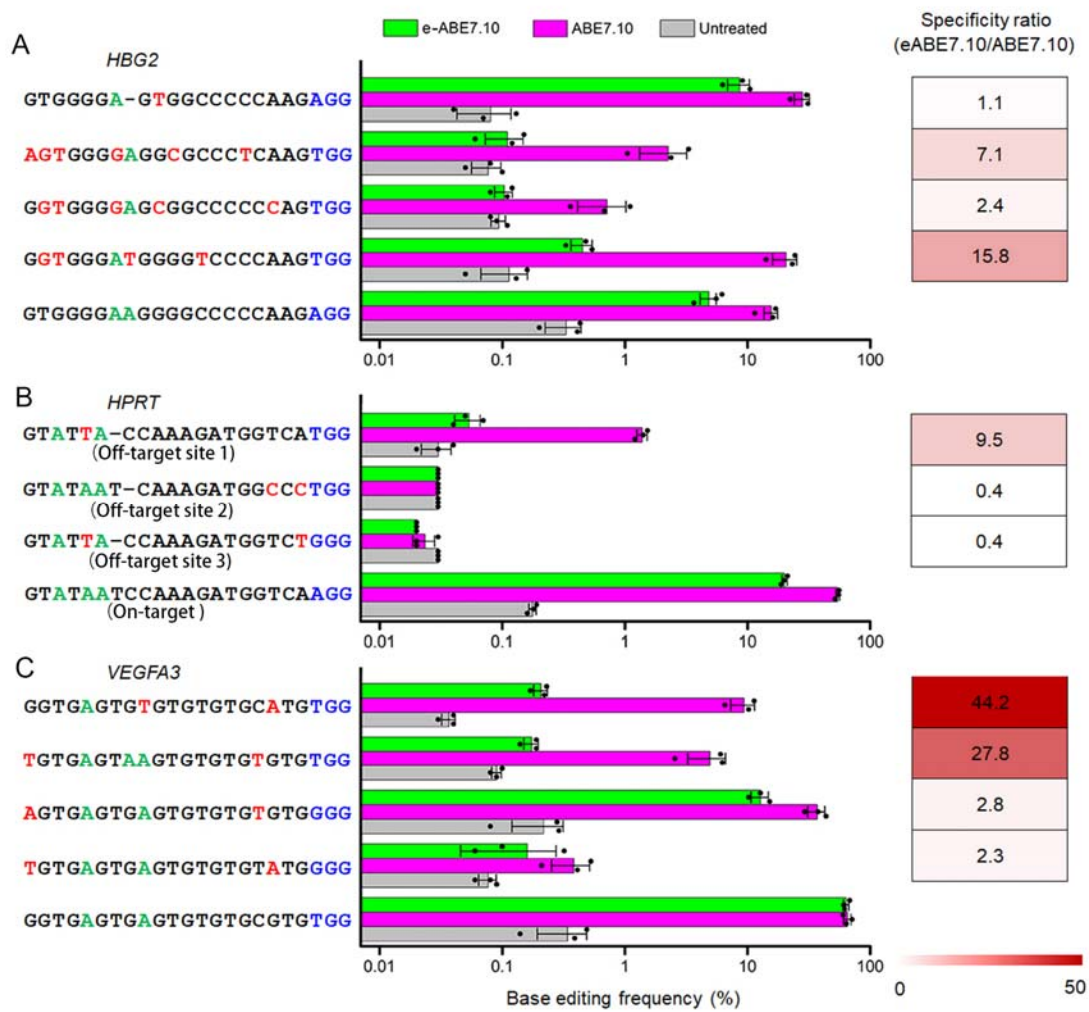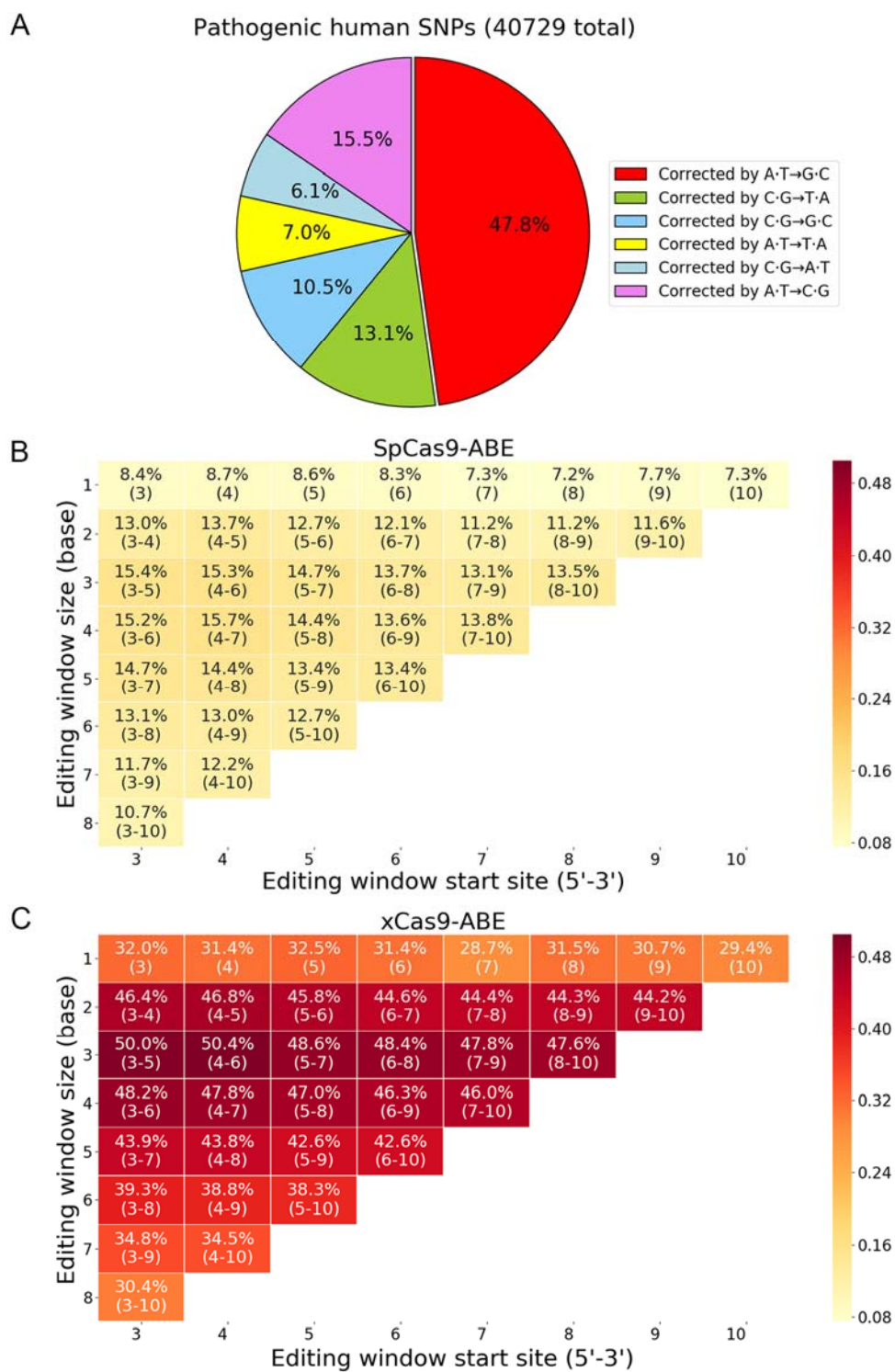
Fig. 1

Fig. 2

Fig. 3