nature biotechnology

Brief Communication

https://doi.org/10.1038/s41587-022-01595-6

Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase

Received: 23 May 2022

Accepted: 1 November 2022

Published online: 09 January 2023

Check for updates

Huawei Tong[1,7] ✉, Xuchen Wang[2,3,7], Yuanhua Liu[2,7], Nana Liu[1,7], Yun Li[1], Jiamin Luo[1], Qian Ma[1], Danni Wu[1], Jiyong Li[1], Chunlong Xu[4] ✉ & Hui Yang[1,2,5,6] ✉

Here we developed an adenine transversion base editor, AYBE, for A-to-C and A-to-T transversion editing in mammalian cells by fusing an adenine base editor (ABE) with hypoxanthine excision protein N-methylpurine DNA glycosylase (MPG). We also engineered AYBE variants enabling targeted editing at genomic loci with higher transversion editing activity (up to 72% for A-to-C or A-to-T editing).

Base editors are promising tools for precise base editing in basic research and therapeutic applications[1,2]. Adenine base editors (ABEs) and cytosine base editors (CBEs) enable A:T to G:C and C:G to T:A transitions, respectively[3,4]. Recently, C-to-G base editors (CGBEs) were developed by replacing uracil glycosylase inhibitor (UGI) with uracil DNA N-glycosylase (UNG) in cytosine base editors[5–9]. However, no editor exists that can enable base conversions including transition and transversion. Base editor enabling A-to-T and A-to-C transversions remains to be achieved to repair a large number of point mutations[2], accounting for up to 27% genetic diseases (Supplementary Fig. 1).

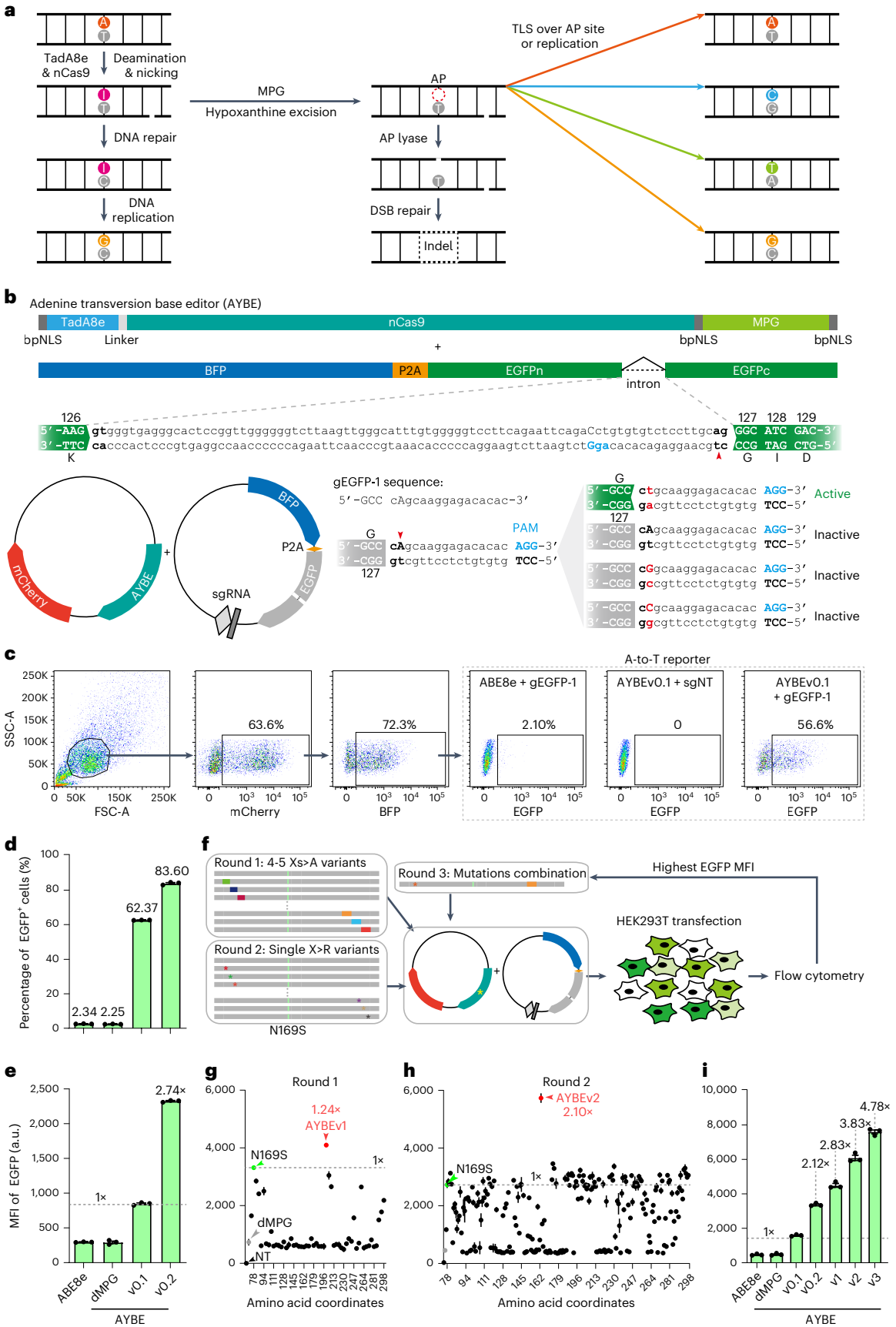
To induce A-to-T and A-to-C transversion editing, we hypothesized that excision of ABE-induced deoxyinosine might enable more versatile base editing outcomes, by triggering the base excision repair (BER) pathway[10,11] in cells (Fig. 1a). We developed three prototype versions of an adenine transversion base editor (AYBE, Y = C or T base) by fusing ABE8e to wild-type human N-methylpurine DNA glycosylase protein (MPG; also known as alkyladenine DNA glycosylase (AAG)), which could excise hypoxanthine (Hx) in damaged DNA[12,13], at different orientations with respect to nCas9 (Fig. 1b and Supplementary Fig. 2a).

To conveniently evaluate the transversion activity of AYBE, we engineered a simple intron-split EGFP reporter system. Disruptive point mutations were introduced in the intron boundary to generate an inactive splicing acceptor signal. A-to-T or A-to-C transversion was required to correct the mutation for proper splicing of EGFP-coding sequence, thus activating EGFP expression (Fig. 1b and Supplementary Fig. 3a). The fluorescence intensity of EGFP could be detected with flow cytometry (Fig. 1c and Supplementary Fig. 3b). After co-transfection with the A-to-T reporter vector, in which the guide RNA (gRNA) targeted the intronic mis-splicing mutation, the AYBE candidate with MPG fused at the C-terminus (TCM, hereafter designated as AYBEv0.1) showed the highest transversion-promoting activity (67.17% versus 63.27% for MTC, 59.03% for TMC) in HEK293T cells (Supplementary Fig. 2b). We readily detected 56.6% and 7.32% of EGFP$^+$ cells using AYBEv0.1 with A-to-T and A-to-C reporter, respectively (Fig. 1c and Supplementary Fig. 3c). By contrast, we observed that ABE8e or AYBE with inactive dead MPG (dMPG, carrying E125A, Y127A and H136A mutations) triggered less than 2.35% of EGFP$^+$ cells (Fig. 1c–e and Supplementary Fig. 3c), probably owing to the endogenous expression of MPG. Moreover, AYBEv0.1 with non-target gRNA could not activate expression of EGFP (Fig. 1c), indicating stringency of reporter and transversion-promoting activity of the catalytic Hx excision domain in AYBE.

To improve AYBE activity, we performed rational mutagenesis of MPG and generated hundreds of AYBE variants for screening, using the A-to-T reporter to evaluate the transversion editing activity. First, we introduced MPG-N169S, a mutation enhancing the Hx excision activity of MPG[14], into AYBEv0.1, thus generating the variant AYBEv0.2. AYBEv0.2 could increase the percentage (up to 83.60%; Fig. 1d) and the mean fluorescence intensity (MFI) (2.74-fold increase; Fig. 1e) of EGFP$^+$ cells compared with AYBEv0.1. We then performed two rounds of mutagenesis and screening based on AYBEv0.2 to further improve

[1]HuiGene Therapeutics Co., Ltd., Shanghai, China. [2]Institute of Neuroscience, State Key Laboratory of Neuroscience, Key Laboratory of Primate Neurobiology, Center for Excellence in Brain Science and Intelligence Technology, Chinese Academy of Sciences, Shanghai, China. [3]College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. [4]Lingang Laboratory, Shanghai, China. [5]Shanghai Research Center for Brain Science and Brain-Inspired Intelligence, Shanghai, China. [6]HuiEdit Therapeutics Co., Ltd., Shanghai, China. [7]These authors contributed equally: Huawei Tong, Xuchen Wang, Yuanhua Liu, Nana Liu. ✉e-mail: huaweitong@huidagene.com; xucl@lglab.ac.cn; huiyang@huidagene.com

https://doi.org/10.1038/s41587-022-01595-6



https://doi.org/10.1038/s41587-022-01595-6

Fig. 1 | Engineering and optimization of AYBE. a, Schematic diagram of potential pathway for adenine transversion and editing outcomes. After adenine deamination by ABE8e and nicking on the non-edited strand by Cas9 nickase (nCas9-D10A), MPG induces Hx excision, followed by DNA repair and/or replication, thus leading to diverse editing outcomes. I, deoxyinosine (the corresponding base is Hx); MPG, N-methylpurine DNA glycosylase; AP, apurinic/apyrimidinic site; DSB, double-strand break. **b**, Schematic designs of reporter and transversion base editor constructs for A-to-T editing detection. Y = C or T. P2A, 2A peptide from porcine teschovirus-1. **c**, Representative flow cytometry scatter plots showing gating strategy and the percentages of EGFP$^+$ cells for each base editor. NT, non-target. **d**, Percentage of EGFP$^+$ cells. **e**, MFI of EGFP. Dotted line, mean value of wild-type MPG group. Fold changes are calculated relative to the wild-type MPG group. a.u., arbitrary units. $n = 3$ in **d**,**e**. **f**, Schematic of mutagenesis and screening strategy. MPG-N169S was a constant mutation during the screening. **g**,**h**, Performance of engineered variants measured by EGFP expression in round 1 and round 2 screening. Each dot represents the mean of three biological replicates of every mutant variant. Dotted line, mean value of the MPG-N169S group. Fold changes are calculated relative to the MPG-N169S group. **i**, Gradual improvement of AYBE-mediated EGFP activation ($n = 3$). Dotted line, mean value of the wild-type MPG group. Fold changes are calculated relative to the wild-type MPG group. All values are presented as mean ± s.e.m.

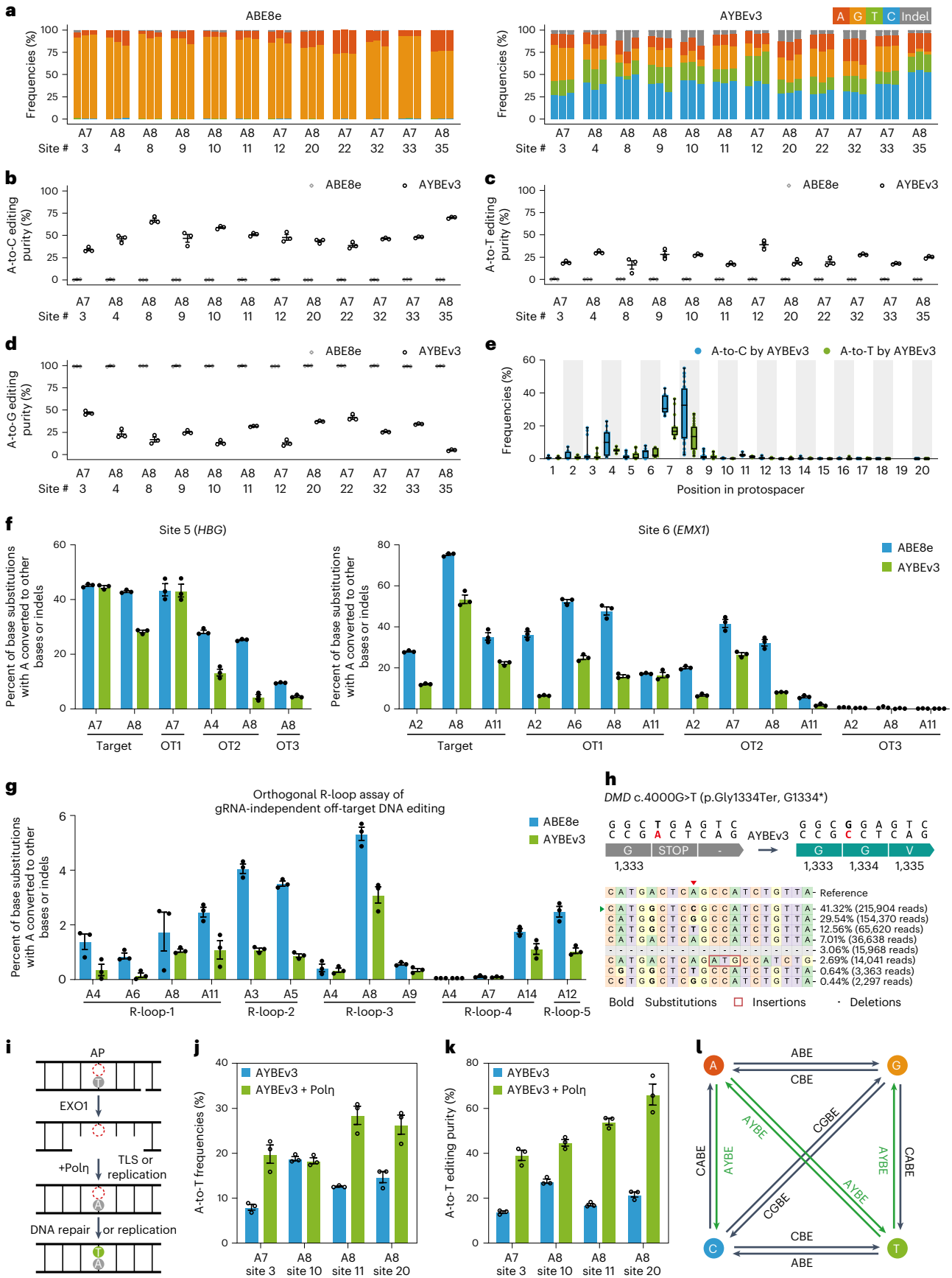
AYBE activity. Based on structural analysis (Supplementary Fig. 2c) and biochemical characterization of MPG, the non-conserved N-terminal region (1–79 amino acids (aa)) has no effect on either base excision or DNA-binding activities of the enzyme[13,15]; the 78–298 aa region of MPG-N169S was evenly divided into 13 segments (F1–F13, 17 aa each) using a recently developed strategy[16]. In the round 1 screening, 52 mutants with four or five random amino acid subsitutions in each segment (replacing all non-alanine to alanine, X > A, and alanine to valine, A > V) distributed near-uniformly in distance were designed and generated, whereas the round 2 mutagenesis scanned the MPG-N169S protein with sequential arginine substitutions (X > R), aiming to enhance the MPG interaction with the substrate DNA (Fig. 1f). Our results showed that most of AYBE variants in round 1 and round 2 screening decreased the transversion editing activity compared with AYBEv0.2, and some variants even lost the activity, similar to AYBE with dMPG (Fig. 1g,h). However, AYBE variant with MPG-F8V1 (termed as AYBEv1, carrying N169S, S198A, K202A, G203A, S206A and K210A) from the round 1 screening and AYBE variant with MPG-G163R and N169S (termed as AYBEv2) from the round 2 screening showed the best performance. AYBEv1 and AYBEv2 exhibited 1.24-fold and 2.10-fold increase of transversion editing activity after normalized to AYBEv0.2 (Fig. 1g,h). To investigate the additive effect of mutations in AYBEv1 and AYBEv2 variants, we combined them in AYBEv3 (carrying G163R, N169S, S198A, K202A, G203A, S206A and K210A) and found synergistic enhancement of transversion editing activity by 4.78-fold in comparison with the prototype version AYBEv0.1 (Fig. 1i). The improvement of transversion editing activity by different AYBE variants from rounds of mutagenesis screening was validated at an endogenous genomic site using fluorescence-activated cell sorting (FACS) (Supplementary Fig. 4a,b). Like the synergistic enhancement of transversion editing activity (from 5.88% to 15.49% for A-to-T editing and from 14.42% to 30.98% for A-to-C editing), we also found synergistic reduction of insertion + deletion (indel) frequencies for AYBEv3 (from 34.28% to 11.64%) (Supplementary Fig. 4c). We speculated that mutations in AYBEv3 might facilitate specific substrate selection or modulate the DNA-binding activity of MPG protein (Supplementary Fig. 4d). Results from the three rounds of mutagenesis screening indicated effective optimization of AYBE toward high activity for A-to-T and A-to-C transversion editing.

We further characterized the editing profiles of AYBEv3 by targeting dozens of endogenous genomic loci. Efficient A-to-C or A-to-T edits were observed with AYBEv3 but almost no A-to-Y (A-to-C or A-to-T) transversion editing at any position of the 26 sites tested with ABE8e (Supplementary Figs. 5–8). The top 12 efficiently edited sites included five sites with an A7 and seven sites with an A8 (Fig. 2a and Supplementary Fig. 5), with A-to-C edits as the predominant product (mean editing frequencies ranging from 34.14% to 70%, up to 70% purity for site 35), with the mean editing frequencies of A-to-T edits ranging from 16.29% to 39.09% (up to 39.09% purity for site 12) (Fig. 2b–d), indicating that AYBEv3 exhibited high editing efficiency for A-to-Y transversion at protospacer positions 7 and 8 (mean editing frequencies ranging from 8% to 72%), including 3–53% editing efficiency for A-to-C transversion and 3–32% for A-to-T transversion (Fig. 2a and Supplementary Fig. 5). Overall, it showed that the editing window of AYBEv3 existed at positions 5–9 on the protospacer and that indels were distributed throughout the protospacer (Supplementary Fig. 9a), with CAA and CAG as the top two preferred edited motifs (Supplementary Fig. 9b). Note that AYBEv3 induced mean indel frequencies (percentage of alleles that contain an insertion or deletion across the entire protospacer) ranging from 1.63% to 40.68% (Supplementary Fig. 9a). In addition, analysis of allele compositions showed that AYBEv3 induced less bystander editing than ABE8e (Supplementary Fig. 10). Moreover, AYBEv3 also exhibited efficient A-to-C and A-to-T transversion editing activity at protospacer positions 7 and 8, with A-to-C edits as the predominant product, across three different human cell lines (HeLa, U2OS and K562 cells) (Supplementary Figs. 11 and 12).

To investigate the off-target (OT) effect of AYBE, we analyzed gRNA-dependent OT activity of AYBEv3 at two previously reported gRNA-dependent OT sites (Fig. 2f) and characterized the ability of AYBEv3 to mediate guide-independent OT DNA editing using orthogonal R-loop assay in five dSaCas9 R-loops[17] (Fig. 2g). We observed a decrease in editing at all six gRNA-dependent OT sites and all five guide-independent OT sites when comparing AYBEv3 to ABE8e (Fig. 2f,g and Supplementary Fig. 13). In addition, we performed a

Fig. 2 | Characterization of editing profiles for AYBE via high-throughput target sequencing. a, Bar plots showing the on-target DNA base editing frequencies of adenines with most A-to-C and/or A-to-T edits with ABE8e and AYBEv3 at the top 12 efficiently edited genomic sites in HEK293T cells. Editing frequencies of three independent replicates ($n = 3$) at each base are displayed side by side. Transfected mCherry$^+$ cells were sorted for further characterization. **b**–**d**, Editing purity of A-to-C (**b**), A-to-T (**c**) or A-to-G (**d**) by ABE8e and AYBEv3 at the edited sites from **a**. **e**, Frequencies of A-to-C and A-to-T editing by AYBEv3 across the protospacer positions 1–20 from the edited sites in **a** (where PAM is at positions 21–23). Single dot represents individual replicate ($n = 3$ independent replicates per site). Boxes span the interquartile range (25th to 75th percentile); horizontal lines indicate the median (50th percentile); and whiskers extend to minima and maxima. **f**, gRNA-dependent OT analysis comparing ABE8e and AYBEv3 at site 5 (*HBG*) and site 6 (*EMX1*) ($n = 3$). Note that a high-fidelity version[17] TadA8e^{V106W} was used in ABE8e and AYBEv3. **g**, gRNA-independent OT editing detected by the orthogonal R-loop assay at each R-loop site ($n = 3$). **h**, Potential correction of *DMD* nonsense mutation by AYBEv3. Allele frequencies of on-target editing by AYBEv3 in stable HEK293T cell lines generated via lentiviral transduction. Arrowheads in red indicate targeted adenines for correction. Arrowheads in green show the allele correction with potential therapeutic benefits. The values in right represent frequencies and reads of mutation alleles. **i**, Schematic diagram of potential pathway to increase the A-to-T editing outcomes. **j**–**k**, A-to-T editing outcomes for the introduction of Polη ($n = 3$). **l**, Diagram showing types of achievable point mutations with the available base editors. All values are presented as mean ± s.e.m.

https://doi.org/10.1038/s41587-022-01595-6



proof-of-concept study to investigate the therapeutic potential of AYBEv3 for correcting disease-related transversion mutations. By testing two nonsense mutations and two splicing acceptor site mutations with AYBEv3 in a stable HEK293T cell line generated via lentiviral transduction, we found approximately 36% and 44% correction frequencies of A-to-C edits at *DMD* and *SLC26A4* nonsense mutation sites and approximately 11% and 20% correction frequencies of A-to-T edits at *ATM* and *TTN* splicing acceptor site mutations, respectively (Fig. 2h and Supplementary Fig. 14), indicating promising potential for AYBE in both basic research and therapeutic applications.

Compared with ABE and CBE, the product purity of AYBE and CGBE needs to be improved for more precise genome editing. In our AYBE-mediated transversion editing process, cellular DNA repair machinery was channeled to favor BER pathway by the activity of Hx excision repair proteins after adenine deamination. We thus attempted to increase the percentage or purity of A-to-T editing by co-expression of AYBEv3 and Polη, a translesion synthesis (TLS) polymerase preferentially incorporating A opposite AP sites[18] (Fig. 2i–k and Supplementary Fig. 15). After co-transfection of plasmids encoding AYBEv3 and Polη, the purity of A-to-T editing outcomes could be substantially increased and achieved up to 66% (Fig. 2k). We have also tested AYBE with a less processive deaminase from ABEmax, termed AYBEmax, and we found that AYBEmax did not lead to more dominant A-to-T or A-to-C editing outcome (Supplementary Fig. 16).

Overall, our findings with engineering a novel AYBE for effective A-to-T and A-to-C editing provide the complementary toolkit to the current base editor repertoire for modeling and treating disease-causing transversion mutations in humans. Henceforth, AYBE, ABE, CBE and CGBE would allow all types of base conversions, including transition and transversion (Fig. 2l). In addition, AYBE could convert A to all other types of bases, thus potentially suited for saturation mutagenesis. AYBEv3 exhibited high editing efficiency for A-to-Y transversion at A7 and A8 while dominantly resulting in A-to-G transitions at other A positions. We speculate that this might result from the activity window or preferred motifs of MPG protein for Hx excision. Meanwhile, it would be necessary to increase transversion editing activity and product purity of AYBE to obtain an A-to-C base editor or an A-to-T base editor via protein engineering or gRNA engineering or to test other DNA repair proteins[10,11,19]. Although the current AYBE version has some shortcomings to correct disease-relevant mutations, we have shown that editing can be shifted from A-to-Y toward A-to-T by co-expression of a TLS polymerase Polη, indicating that it is potentially possible to engineer both ATBE and ACBE out of AYBE, which would markedly increase the therapeutic potential of the base editor platform. Moreover, there are other edited adenines in the window that cause mutations (Fig. 2h and Supplementary Fig. 14d–f), and a more accurate AYBE with a refined editing window could bypass this issue in the future.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-022-01595-6.

References

1. Porto, E. M., Komor, A. C., Slaymaker, I. M. & Yeo, G. W. Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.* **19**, 839–859 (2020).
2. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
3. Gaudelli, N. M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
4. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
5. Zhao, D. et al. Glycosylase base editors enable C-to-A and C-to-G base changes. *Nat. Biotechnol.* **39**, 35–40 (2021).
6. Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).
7. Koblan, L. W. et al. Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat. Biotechnol.* **39**, 1414–1425 (2021).
8. Chen, L. et al. Programmable C:G to G:C genome editing with CRISPR–Cas9-directed base excision repair proteins. *Nat. Commun.* **12**, 1384 (2021).
9. Yuan, T. et al. Optimization of C-to-G base editors with sequence context preference predictable by machine learning methods. *Nat. Commun.* **12**, 4902 (2021).
10. Robertson, A. B., Klungland, A., Rognes, T. & Leiros, I. DNA repair in mammalian cells: base excision repair: the long and short of it. *Cell. Mol. Life Sci.* **66**, 981–993 (2009).
11. Hindi, N. N., Elsakrmy, N. & Ramotar, D. The base excision repair process: comparison between higher and lower eukaryotes. *Cell. Mol. Life Sci.* **78**, 7943–7965 (2021).
12. Saparbaev, M. & Laval, J. Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl Acad. Sci. USA* **91**, 5873–5877 (1994).
13. Lau, A. Y., Scharer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258 (1998).
14. Connor, E. E. & Wyatt, M. D. Active-site clashes prevent the human 3-methyladenine DNA glycosylase from improperly removing bases. *Chem. Biol.* **9**, 1033–1041 (2002).
15. Vallur, A. C., Maher, R. L. & Bloom, L. B. The efficiency of hypoxanthine excision by alkyladenine DNA glycosylase is altered by changes in nearest neighbor bases. *DNA Repair (Amst).* **4**, 1088–1098 (2005).
16. Tong, H. et al. High-fidelity Cas13 variants for targeted RNA degradation with minimal collateral effects. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01419-7 (2022).
17. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
18. Choi, J. Y., Lim, S., Kim, E. J., Jo, A. & Guengerich, F. P. Translesion synthesis across abasic lesions by human B-family and Y-family DNA polymerases α, δ, η, ι, κ, and REV1. *J. Mol. Biol.* **404**, 34–44 (2010).
19. Thompson, P. S. & Cortez, D. New insights into abasic site repair and tolerance. *DNA Repair (Amst).* **90**, 102866 (2020).

https://doi.org/10.1038/s41587-022-01595-6

Methods

Molecular cloning

Base editor constructs used in this study were cloned into a mammalian expression plasmid backbone under the control of an EF1α promoter by standard molecular cloning techniques. KOD-Plus-Neo DNA polymerase (KOD-401, Toyobo) was used to amplify the insertion fragments, and NEBuilder HiFi DNA Assembly Master Mix (E2621L, New England Biolabs) was used to perform the Gibson assembly of multiple DNA fragments. The Gibson reaction was then transformed into chemically competent *Escherichia coli* DH5α.

The wild-type MPG sequence (298 aa long) was PCR-amplified from cDNA of HEK293T and fused to ABE8e at three different orientations with respect to nCas9 via the Gibson assembly method. Thus, bpNLS-MPG-Linker-TadA8e-Linker-nCas9(D10A)-bpNLS, bpNLS-TadA8e-Linker-MPG-Linker-nCas9(D10A)-bpNLS and bpNLS-TadA8e-Linker-nCas9(D10A)-bpNLS-MPG-bpNLS fusion proteins were generated as initial versions of AYBE for A-to-T and A-to-C editing. MPG-N169S was introduced via site-directed mutagenesis by PCR. The amino acid sequence for AYBEv3 is supplied in Supplementary Table 1.

To improve the transversion activity of AYBE, we developed a simple and convenient reporter system. The reporter BFP-P2A-EGFP driven by a CAG promoter and the U6-gRNA-scaffold were constructed in one single vector. Intron-split EGFP reporters were engineered by insertion of the last intron (86 base pairs (bp) long) of human *RPS5* between the K126 and G127 codons of EGFP. Modification of the 68th base (G > C) or the 70th base (T > C) in the intron sequence for introducing artificial protospacer adjacent motif (PAM) on the template strand, and corresponding mutations at the splice acceptor site, were made to construct A-to-T reporter or A-to-C reporter via site-directed mutagenesis by PCR, respectively. Mutations at the splice acceptor site led to inactive EGFP production by non-spliced EGFP transcripts. Transversion corrections in A-to-T reporter or A-to-C reporter were required for proper splicing of EGFP-coding sequence. Correctly spliced EGFP transcript could produce active EGFP. The gRNA oligos were annealed and ligated into BpiI sites.

For disease-related single nucleotide variant (SNV) transversion editing, four disease-related mutations with the upstream and downstream flanking sequences (50 bp) were constructed in tandem into lentivirus vector. The human Polη sequence was PCR-amplified from cDNA of HEK293T. bpNLS-Polη-P2A-BFP driven by a CAG promoter was constructed by standard molecular cloning techniques.

Design and construct of MPG mutants

MPG-N169S was a constant mutation during the screening. MPG mutagenesis libraries were designed and generated as previously described[16]. MPG-N169S (78–298 aa) was divided into 13 segments, with each 17 aa long. Thirteen BpiI-harboring mutants were introduced via site-directed mutagenesis by PCR. In the round 1 screening, 52 mutants were designed, with four or five random mutation sites distributed near-uniformly in distance for each variant. All non-alanine amino acids were replaced with alanine (X > A). To cover all the residues in the segments mentioned herein, we also mutated alanine to valine (A > V). In the round 2 screening, 221 mutants scanning the protein with sequential arginine (X > R) substitutions were designed, with all arginine amino acids replaced with lysine to cover all the residues in the segments mentioned here, because both have similar size and charge. Oligos coding for mutants in the two rounds of screening were annealed and ligated into corresponding BpiI-digested backbone vectors. The MPG mutants and corresponding codon substitutions used are listed in Supplementary Tables 2 and 3.

Cell culture, transfection and flow cytometry analysis

HEK293T, Hela and U2OS cells were cultured with DMEM (11995065, Gibco) supplemented with 10% FBS (04-001-1ACS, BI Worldwide) and 0.1 mM non-essential amino acids (11140-050, Gibco). K562 cells were cultured with RPMI-1640 (11875-093, Gibco) supplemented with 10% FBS (04-001-1ACS, BI Worldwide), 1% penicillin–streptomycin (15070-063, Gibco) and 0.1 mM non-essential amino acids (11140-050, Gibco). Cells were grown in an incubator at 37 °C with 5% CO_2.

MPG mutant screening was conducted in 48-well plates or 24-well plates. The day before transfection, 3×10^4 HEK293T cells per well were plated in 250 µl of complete growth medium in the 48-well plates. After 12 h, 100 ng of AYBE plasmids and 200 ng of A-to-T reporter plasmids were co-transfected into cells with 600 ng of polyethylenimine (PEI) (DNA:PEI ratio of 1:2.5) per well. In the 24-well plates, 2×10^5 cells were plated per well in 500 µl of complete growth medium, and 150 ng of AYBE plasmids and 300 ng of reporter plasmids were co-transfected into HEK293T cells with 900 ng of PEI.

Disease-related SNV transversion editing was tested in stable HEK293T cell lines via lentiviral. For lentivirus packaging, plasmid with disease-related mutations (1.2 µg) was co-transfected with the packaging plasmids Pax2 (0.9 µg) and Vsvg (0.6 µg) into HEK293T cells using the FuGENE HD transfection reagent (E2311, Promega). After 72 h, lentivirus-containing media was collected for infection and then filtered through a 0.45-µm low protein binding membrane (Millipore). For lentiviral infection, HEK293T cells were dissociated by trypsin-EDTA (25200-072, Gibco), and suspensions were diluted to 18×10^5 cells per well in six-well plates and incubated with 150 µl of lentiviruses for 48 h. Then, the medium was replaced with fresh complete medium.

For cell transfection of HEK293T, Hela, U2OS and K562 cells for FACS, 5×10^5 cells per well were plated in 12-well plates with 1 ml of complete growth medium the day before transfection. After 14–16 h, 2 µg of AYBE-gRNA plasmids were transfected into cells using PEI (DNA:PEI ratio of 1:2.5) or FuGENE HD transfection reagent (E2311, Promega) (DNA:FuGENE ratio of 1:3).

Orthogonal R-loop assays were performed as described previously[17], with minor modifications. Then, 1 µg of AYBE plasmid with single guide RNA (sgRNA) targeting site 3 and 1 µg of dSaCas9 plasmid with corresponding sgRNA targeting five OT sites to generate R-loops were co-transfected into HEK293T cells in 12-well plates using PEI (DNA:PEI ratio of 1:2.5).

Forty-eight hours after transfection, expression of mCherry, BFP and EGFP fluorescence was analyzed by BD FACSAria III or Beckman CytoFLEX S. Flow cytometry results were analyzed with FlowJo version 10.5.3. The gating strategy in the identification of $mCherry^+$, BFP^+ and $EGFP^+$ cells for on-target editing efficiency evaluation is supplied in Fig. 1c.

Target sequencing of endogenous sites

At 72 h after transfection, 10,000 $mCherry^+$ cells were isolated by FACS. gDNA was extracted by the addition of 40 µl of lysis buffer and 1 µl of proteinase K (PD101-01, Vazyme) directly into each tube of sorted cells. The gDNA/lysis buffer mixture was incubated at 55 °C for 45 min, followed by a 95 °C enzyme inactivation step for 10 min. The regions of interest for target sites were amplified by PCR using site-specific primers. The PCR reaction was performed at 95 °C for 5 min, 28 cycles at 95 °C for 15 s, 60 °C for 15 s, 72 °C for 30 s and a final extension at 72 °C for 5 min using Phanta Max Super-Fidelity DNA Polymerase (P505-d3, Vazyme). PCR products were purified using universal DNA purification kit (TIANGEN) according to the manufacturer's instructions and analyzed by Sanger sequencing (GENEWIZ). The amplicons were ligated to adapters, and sequencing was performed on the Illumina MiSeq platform. Protospacer sequences and site-specific primers used for each genomic locus are listed in Supplementary Tables 4 and 5.

Target sequencing data analysis

Targeted amplicon sequencing reads were first input to trim_galore (powered by Cutadapt 0.6.6) for quality trimming, and the reads with fewer than 30 bp were filtered. The cleaned pairs were then merged using FLASH version 1.2.11. The amplified sequences from individual targets were demultiplexed using fastx_barcode_splitter.pl from the

fastx_toolkit (0.0.14). Further amplicon sequencing analysis was performed by CRISPResso2 (ref. 20). A 10-bp window was used to quantify modifications centered around the middle of the 20-bp gRNA. Otherwise, the default parameters were used for analysis. The output files, 'Quantification_window_nucleotide_frequency_table.txt' and 'Quantification_window_modification_count_vectors.txt', were combined to calculate the base substitution and indel rates for each individual targeting. In brief, counts of nucleotide bases (A, C, G and T) as well as deletion (−) and ambiguous bases (N) for each position in sgRNA were extracted from 'alleles_frequency_table_around_sgRNA_*.txt'. The aligned sequences with inserted bases were assigned to the reference base when insertions appear for some specific position. To give a global view of the modifications of individual position of the reference, the counts of the insertions from 'Quantification_window_modification_count_vectors.txt' were introduced and used to verify the counts of the reference base though subtracting the insertion counts from the counts of reference base. The verified counts of the nucleotide bases (A, C, G and T) as well as indels were further used to calculate the base substitution and indel rates for each position of sgRNA.

Statistical analysis

Statistical tests performed by GraphPad Prism 8 included the two-tailed, unpaired, two-sample *t*-test or Dunnett's multiple comparisons test after one-way ANOVA. All values are reported as mean ± s.e.m.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Expression plasmids used in this study have been deposited at Addgene and will be available at https://www.addgene.org/Huawei_Tong/ (Addgene plasmid nos. 193966–193968). All data supporting the findings of this study are available in the paper (and in its Supplementary Information files). Targeted amplicon sequencing data have been deposited at the Sequence Read Archive and can be accessed at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA874457 (ref. 21). All relevant original data are available from the corresponding authors upon reasonable request.

Code availability

Custom scripts for CRISPResso analyses supporting the findings of this study are available from the corresponding author upon reasonable request.

References

20. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
21. Tong, H. et al. Sequence Read Archive. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA874457 (2022).

Acknowledgements

We acknowledge technical support from the FACS facility of HuiGene Therapeutics Co., Ltd. This work was supported by HuiGene Therapeutics Co., Ltd. (H.T.).

Author contributions

H.Y. and H.T. jointly conceived the project. H.T. designed and conducted experiments. X.W. and N.L. performed experiments. Y.L. performed target sequencing data analysis. Y.L. and J.L assisted with cell experiments. Q.M., D.W. and J.L. participated in vector construction and FACS. H.Y., H.T. and C.X. supervised the whole project. H.Y., C.X. and H.T. wrote the manuscript draft, and all authors contributed to the editing of the manuscript.

Competing interests

H.T. discloses a patent application related to the proteins described in this manuscript. H.T., N.L., Y.L., J.L, Q.M., D.W. and J.L. are employees of HuiGene Therapeutics Co., Ltd. H.Y. is a founder of HuiGene Therapeutics Co., Ltd. and HuiEdit Therapeutics Co., Ltd. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-022-01595-6.

Correspondence and requests for materials should be addressed to Huawei Tong, Chunlong Xu or Hui Yang.

Peer review information *Nature Biotechnology* thanks the anonymous reviewers for their contribution to the peer review of this work.

nature research

Corresponding author(s): Hui Yang

Last updated by author(s): Oct 31, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Next-generation sequencing data were obtained on the Illumina MiSeq platforms.
Data analysis	FlowJo V10.5.3, Graphpad Prism(v8.3.0), Cutadapt(v0.6.6), FLASH (v1.2.11), fastx_toolkit (v0.0.14), CRISPResso2(v2.0.20b); Custom scripts for CRISPResso analyses supporting the findings of this study are available from the corresponding author upon reasonable request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data supporting the findings of this study are available in the paper (and in its supplementary information files). Targeted amplicon sequencing data have been deposited at the Sequence Read Archive and can be accessed via: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA874457. Any relevant original data are available from the corresponding authors upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculation was performed. Experiments were performed in triplicates(n=3), unless otherwise moted, with the sample size numbers were listed in the corresponding figure legends. Sample sizes for these experiments were chosen based upon field standards and prior knowledge of experimental variation. (e.g. Gaudelli et al., Nature 2017; Koblan et al., Nat Biotechnol 2021)
Data exclusions	No data were excluded.
Replication	Data were obtained in the fashion of biological replications. We tested experimental conditions using different gRNAs to ensure robustness. All the experimental results could be successfully reproduced. At least three independent experiments were performed.
Randomization	For transfection, wells were randomly assigned.
Blinding	Blinding was not relevant to our study because it is not a subjective trial and NGS and other methods used for quantification in our study are not influenced by human interpretation and/or bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T, Hela, U2OS cell lines were purchased from Stem Cell Bank, Chinese Academy of Sciences; K562 cell lines were purchased from BNCC.
Authentication	Cell lines were authenticated by the supplier.
Mycoplasma contamination	Cell lines have been tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell line listed in the database of ICLAC was used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	To isolate cells, the transfected or non-transfected were dissociated enzymatically in an incubation solution of 50 µL Trypsin-EDTA (0.05%) at 37°C for 5 min. The digestion was stopped by adding 500 µL of DMEM medium with 10% Fetal Bovine Serum (FBS). The cell suspension was centrifuged for 3 min (1000 rpm), and the pellet was resuspended in 300 µL DMEM medium with 10% FBS. Finally, the cell suspension was filtered through a 40-µm cell strainer, and mCherry+ cells were isolated by FACS.
Instrument	BD FACS Aria III, Beckman CytoFLEX S
Software	FlowJo V10.5.3
Cell population abundance	Samples were found to be >95% pure when assessed with a second round of flow cytometry analysis.
Gating strategy	Gating strategy: 1) in FSC-A/SSC-A or FSC-H/SSC-H gate for living cells, 2) using the non-transfected and negative cells to define the gates for BFP+, EGFP+ and/or mCherry+ cells, 3) apply this gate to all samples.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.