nature biotechnology

BRIEF COMMUNICATION

https://doi.org/10.1038/s41587-021-00868-w

High-efficiency prime editing with optimized, paired pegRNAs in plants

Qiupeng Lin[1,2,6], Shuai Jin[1,2,6], Yuan Zong[1,6], Hong Yu[3,6], Zixu Zhu[1,2], Guanwen Liu[4,2], Liquan Kou[3], Yanpeng Wang[1,2], Jin-Long Qiu[4,5], Jiayang Li[2,3] ✉ and Caixia Gao[1,2] ✉

Prime editing (PE) applications are limited by low editing efficiency. Here we show that designing prime binding sites with a melting temperature of 30 °C leads to optimal performance in rice and that using two prime editing guide (peg) RNAs *in trans* encoding the same edits substantially enhances PE efficiency. Together, these approaches boost PE efficiency from 2.9-fold to 17.4-fold. Optimal pegRNAs or pegRNA pairs can be designed with our web application, PlantPegDesigner.

Introducing desired mutations into plant genomes could provide considerable economic benefits for agriculture[1,2]. The newly developed prime editors can generate all types of base conversions and also small insertions and deletions[3–13]. PE uses CRISPR–Cas9 nickase (H840A)-reverse transcriptase (RT) fusion proteins programmed with pegRNAs that also contain a prime binding site (PBS) sequence and an RT template. The PBS sequence hybridizes with the 3′ end of the protospacer adjacent motif (PAM)-containing strand and is nicked by the PE–pegRNA complex. The desired edits in the RT template are then reverse transcribed and inserted into the target site[3]. However, the efficiency of prime editors is low and needs improving, especially in the case of low-efficiency targets[4–10]. Previous studies demonstrated that the efficiency of prime editors is strongly affected by the design of the pegRNA[3,4,12], but the reason for this is unclear. Here we show that PE efficiency can be improved by optimizing the melting temperature (T_m) of the PBS and by using a dual-pegRNA strategy in plants. We also describe a web application, PlantPegDesigner, that simplifies the design of pegRNAs based on these principles.

Because T_m is important for the stability of DNA, RNA and DNA/RNA duplexes[14], we hypothesized that the T_m of the PBS sequence (referred to as PBS T_m) is an important parameter for plant prime editors (PPEs)[4]. Previous studies[4,5,8–11] showed that the plant prime editor PPE2 has similar editing efficiency in plants to PPE3, whereas the latter needs a second single guide RNA to nick the non-edited strand. Therefore, in this study, we used PPE2 (referred to hereafter as PPE) to evaluate editing efficiency. We first assessed the effects of T_m-controlling PBS length on PE in plants based on published data[4] (Fig. 1a). The results indicated that PPEs were most efficient when PBS T_m approached 30 °C (30 °C in *OsCDC48-T1*, 28 °C in *OsCDC48-T2* and 30 °C in *OsALS-T1*) (Fig. 1a). We, therefore, evaluated the editing efficiencies of PPE at four targets—*OsACC-T1*, *OsEPSPS-T1*, *OsCDC48-T3* and *OsPDS-T1*—in rice protoplasts using PBS T_m ranging from 18 °C to 52 °C (corresponding to PBS lengths ranging from 6 nt to 17 nt) (Fig. 1b and Supplementary Table 1). We found that pegRNAs had much higher activities when the PBS T_m was around 30 °C (24–30 °C in *OsACC-T1*, 26–34 °C in *OsEPSPS-T1*, 28–36 °C in *OsCDC48-T3* and 30 °C in *OsPDS-T1*) (Fig. 1b); at these PBS T_m temperatures, they were 1.5-fold to 4.3-fold higher than at other PBS T_m temperatures. When we tested 11 more targets, we found that nine of the 11 targets (the exceptions being *OsEPSPS-T2* and *OsNRT1.1B-T1*) behaved in the same way (Fig. 1b). We then normalized and compared the overall editing efficiency of PPEs at different PBS T_m temperatures at all 18 targets. The results indicated that PE efficiencies obeyed a normal distribution ($P > 0.1$) (Fig. 1c) and generally were maximal at PBS T_m 30 °C, followed by PBS T_m 32 °C and 28 °C, and decreased on either side of these PBS T_m temperatures (Fig. 1c and Supplementary Fig. 1). We conclude that the T_m of the PBS sequence is strongly related to PPE editing efficiency (Fig. 1c) and is likely to be a major factor influencing the design of plant pegRNAs. Thus, we recommend using a 30 °C PBS T_m with PPE.

To optimize PE, we developed a dual-pegRNA strategy using separate pegRNAs in trans encoding the same edits for the forward and reverse for each of the DNA strands simultaneously (referred to as NGG-pegRNA and CCN-pegRNA, respectively) (Fig. 1d). We chose 15 targets in nine rice genes and designed a dual-pegRNA for each (Supplementary Table 2). We then compared the PE activities of the NGG-pegRNA only, the CCN-pegRNA only and the dual-pegRNA at the same sites. We found that the dual-pegRNA strategy gave the highest activities at most of the targets (13 of 15); they generated C-to-A, G-to-A, G-to-T, A-to-G, T-to-A, C-to-G and CT-to-AG point mutations, 1-bp (T) or 2-bp (AT) deletion and 1-bp (A) insertion, with maximal editing efficiencies reaching 24.5% (Fig. 1e). The PE efficiency of the dual-pegRNA for all tested sites was about 4.2-fold higher (maximum 27.9-fold for *OsNRT1.1B* (A ins)) than that created by the individual NGG-pegRNAs and 1.8-fold on average (maximum 7.2-fold for *OsALS* (A-to-G)) higher than that produced by the individual CCN-pegRNAs (Fig. 1e,f). In particular, PE by the dual-pegRNA strategy was greatly improved at *OsNRT1.1B* (A ins), *OsNRT1.1B* (G-to-A) and *OsODEV* (CT-to-AG) (up to 2.7%), at which the NGG-pegRNAs produced virtually no corresponding editing events (Fig. 1e). Note that the proportion of byproducts, including undesired indels, was not higher using the dual-pegRNAs than the single pegRNAs (Fig. 1e and Supplementary Fig. 2).

[1]State Key Laboratory of Plant Cell and Chromosome Engineering, Center for Genome Editing, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China. [2]College of Advanced Agricultural Sciences, University of Chinese Academy of Sciences, Beijing, China. [3]State Key Laboratory of Plant Genomics, and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing, China. [4]State Key Laboratory of Plant Genomics, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China. [5]CAS Center for Excellence in Biotic Interactions, University of Chinese Academy of Sciences, Beijing, China. [6]These authors contributed equally: Qiupeng Lin, Shuai Jin, Yuan Zong, Hong Yu. ✉e-mail: jyli@genetics.ac.cn; cxgao@genetics.ac.cn

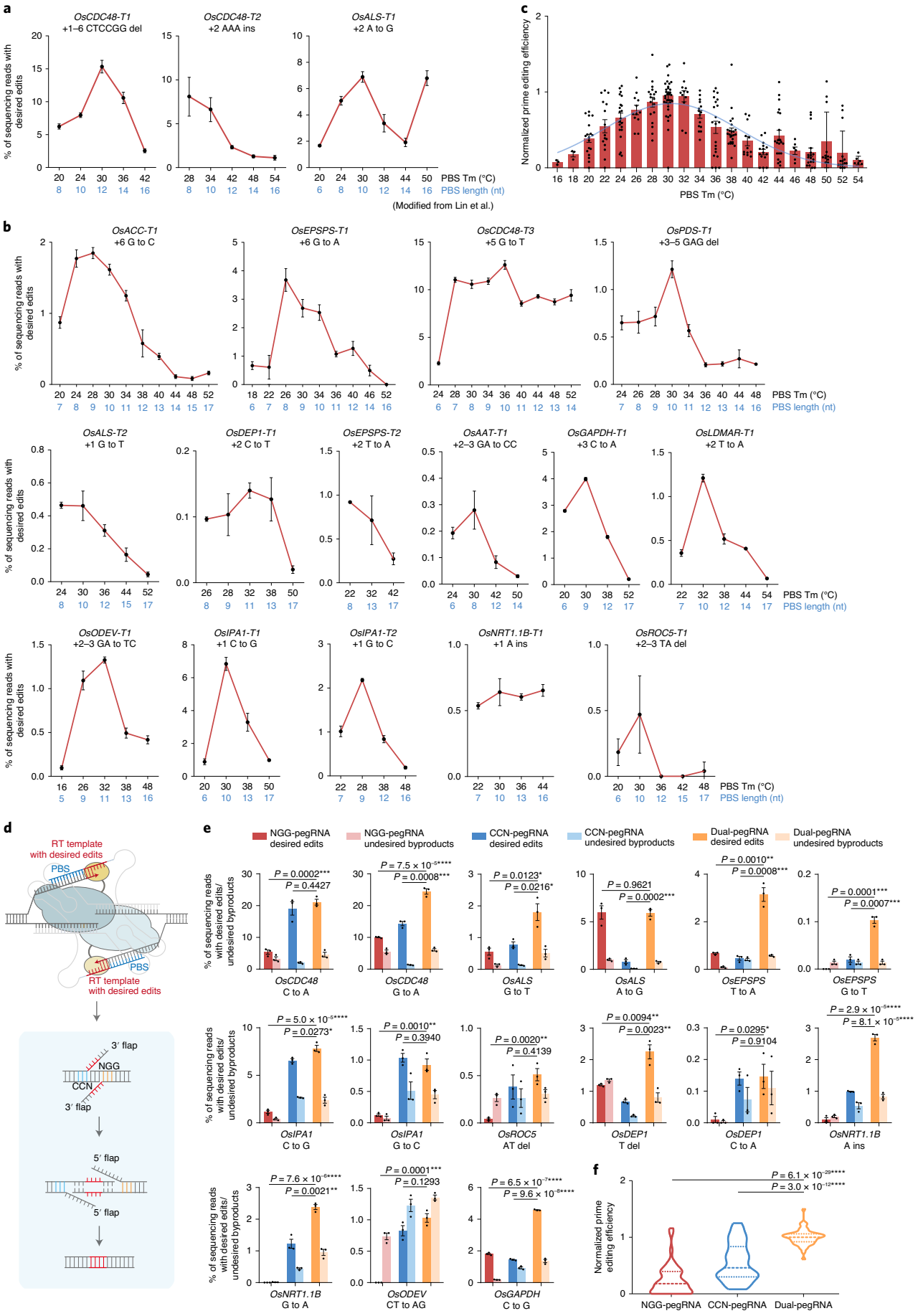
a
OsCDC48-T1
+1–6 CTCCGG del
OsCDC48-T2
+2 AAA ins
OsALS-T1
+2 A to G
% of sequencing reads with desired edits
PBS Tm (°C)
PBS length (nt)
(Modified from Lin et al.)
c
Normalized prime editing efficiency
PBS Tm (°C)
b
OsACC-T1
+6 G to C
OsEPSPS-T1
+6 G to A
OsCDC48-T3
+5 G to T
OsPDS-T1
+3–5 GAG del
OsALS-T2
+1 G to T
OsDEP1-T1
+2 C to T
OsEPSPS-T2
+2 T to A
OsAAT-T1
+2–3 GA to CC
OsGAPDH-T1
+3 C to A
OsLDMAR-T1
+2 T to A
OsODEV-T1
+2–3 GA to TC
OsIPA1-T1
+1 C to G
OsIPA1-T2
+1 G to C
OsNRT1.1B-T1
+1 A ins
OsROC5-T1
+2–3 TA del
d
RT template with desired edits
PBS
PBS
RT template with desired edits
3′ flap
NGG
CCN
3′ flap
5′ flap
5′ flap
e
NGG-pegRNA desired edits
NGG-pegRNA undesired byproducts
CCN-pegRNA desired edits
CCN-pegRNA undesired byproducts
Dual-pegRNA desired edits
Dual-pegRNA undesired byproducts
% of sequencing reads with desired edits/ undesired byproducts
OsCDC48 C to A
OsCDC48 G to A
OsALS G to T
OsALS A to G
OsEPSPS T to A
OsEPSPS G to T
OsIPA1 C to G
OsIPA1 G to C
OsROC5 AT del
OsDEP1 T del
OsDEP1 C to A
OsNRT1.1B A ins
OsNRT1.1B G to A
OsODEV CT to AG
OsGAPDH C to G
f
Normalized prime editing efficiency
NGG-pegRNA
CCN-pegRNA
Dual-pegRNA

Fig. 1 | Comparison of the effects on PE in rice protoplasts of varying T_m-directed PBS lengths and forms of pegRNA. a, **b**, Comparison of the effect on PE of varying T_m-directed PBS lengths at three target sites using published data[4] (**a**) and at 15 new target sites using newly obtained data (**b**) in rice protoplasts. **c**, Normalized PE frequencies with different PBS melting temperatures. The average editing efficiencies of three repeats of the highest editing efficiency obtained at each target was normalized to 1, and the frequencies obtained at the other PBS T_m were adjusted accordingly (18 target sites, $n = 3$ independent experiments). The red column represents the average of the normalized editing efficiencies. The efficiencies were found to follow a normal distribution (shown by the blue line) using the Kolmogorov–Smirnov test. The fitted curve equation is $y = 0.8523 \times \exp(-0.5 \times ((x - 30.37)/8.415)^2)$ ($R^2 = 0.8101$). **d**, Diagram of PE using the dual-pegRNA strategy. The 3′ flap containing the desired edits is reverse transcribed by the prime editor. Equilibration between the edited 3′ flap and the unedited 5′ flap, followed by DNA repair, result in stably edited DNA. The dual-pegRNA creates the edits simultaneously in both DNA strands. **e**, Comparison of the efficiencies of PPEs induced by the NGG-pegRNA, CCN-pegRNA and dual-pegRNA strategies at 15 target sites. The edits were referred to the base on the DNA forward strand. **f**, Overall editing frequencies induced by PPEs containing NGG-pegRNA, CCN-pegRNA and dual-pegRNA. The editing frequencies using dual-pegRNAs for each target were normalized to 1, and the frequencies using NGG-pegRNA and CCN-pegRNA for each target were adjusted accordingly (15 target sites, $n = 3$ independent experiments). Frequencies (mean ± s.e.m.) in **a** and **b** and **e** and **f** were calculated from three independent experiments ($n = 3$). *P* values were obtained using two-tailed Student's *t*-tests in **e** and **f**. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ and $^{****}P < 0.0001$.

We also sought to expand the targeting scope of PPEs and further increase the utility of dual-pegRNAs using the SpG, variant of SpCas9 (ref. [15]) (Supplementary Table 3). We found that SpG-PPE had a relatively broad ability to prime edit NG PAM sequences, with efficiencies of up to 1.9% (Supplementary Fig. 3a). In addition, we examined editing efficiencies at another two NG PAM target sites using the dual-pegRNA strategy. Not surprisingly, editing by the dual-pegRNAs was much more efficient than by their individual component pegRNAs (Supplementary Fig. 3b). These findings show that the dual-pegRNA strategy markedly increases PE activity.

Computational analysis of the rice reference genome (Os-Nipponbare reference IRGSP-1.0) revealed that dual-pegRNA could theoretically target 21.5% of genomic bases when the PE window was defined as extending from +1 to +15 and the canonical SpCas9 was used (Supplementary Fig. 4). However, when combined with the SpG variant with an NG PAM, the dual-pegRNA strategy could potentially target 89.2% of the rice bases (Supplementary Fig. 4).

Although there are some web applications that aim to simplify the design of pegRNAs, their design strategies are based on general pegRNA design principles and require potential pegRNAs to be tested experimentally[16–19]. In this study, we developed PlantPegDesigner, a user-friendly web application (http://www.plantgenomeediting.net/) based on the principles detailed above: optimizing T_m-directed PBS length design, using dual-pegRNA and excluding the first C in the 3′ extension of the RT template (Supplementary Fig. 5 and Supplementary Table 4), together with some other published general design principles concerning RT template length, PBS GC content and PE window[3,20] (see details in Supplementary Note 1). PlantPegDesigner provides precise guidance on details of any intented PE experiment in plants, making recommendations concerning the on-target spacer, PBS sequence, RT template sequence and primers for vector construction (Fig. 2a).

For any given PPE experiment, PlantPegDesigner needs only a single input sequence, including the reference and edited sequence (Supplementary Fig. 6a and see details in Supplementary Note 2). PlantPegDesigner provides a variety of choices of parameters to meet the different needs of users (see details in Supplementary Note 1).

PlantPegDesigner first screens the forward/reverse strand of the input sequence of the spacer sequence and PAM (spacer-PAM) to check if the desired edits are correctly positioned in the user-defined PE window (Fig. 2a and Supplementary Fig. 6a). The dual-pegRNA model will then be recommended if the spacer-PAM sequences can be found in both the forward and reverse strands (Fig. 2a). PlantPegDesigner displays all possible candidate PBS and RT template sequences of varying length (Fig. 2a and Supplementary Fig. 6b). It then recommends the PBS sequences with user-defined optimal PBS T_m (default to 30 °C, based on experimental data for T_m-directed PBS length) and an RT template sequence based on previously published design principles[3,17] (Supplementary Fig. 6b and see details in Supplementary Note 2). For each pegRNA, PlantPegDesigner designs the primer sets for a one-step polymerase chain reaction (PCR) strategy for vector construction and also supports the batch design for multiple input sequences (see details in Supplementary Notes 3 and 4).

We used previously published data to validate the PlantPegDesigner design algorithm with default parameters[4]. We found that PlantPegDesigner performed well along with recommended PBS T_m and RT template length at three previously published, randomly selected rice sites (Supplementary Fig. 7). We also used PlantPegDesigner to design eight different pegRNAs and compared their editing efficiencies in rice protoplasts with the manual-designed pegRNAs with ~13-nt PBS sequences created by following previously published general design guidelines[3] (Fig. 2b). The results indicated that the algorithm-designed pegRNAs had markedly higher editing efficiencies than the manual-designed ones across five targets and three targets using dual-pegRNA (Fig. 2b,c and Supplementary Table 5). To further test the design algorithm experimentally, we compared the PE efficiencies of pegRNAs designed with PlantPegDesigner and those designed with three other web applications: multicrispr[16], pegFinder[17] and PrimeDesign[18] (Supplementary Table 6). We found that pegRNAs designed by PlantPegDesigner had higher editing efficiencies at each of the sites tested than the other pegRNAs (Fig. 2d). Notably, the PlantPegDesigner-recommended dual-pegRNAs were much more efficient than the pegRNAs designed by the other web applications when editing the same edits; the data in Fig. 2e show that they were, on average, 1.8-, 16.3- and 45.8-fold higher at *OsCDC48*, *OsNRT1.1B* and *OsALS* targets. These results provided experimental confirmation of the usefulness of PlantPegDesigner.

In conclusion, we showed that a design strategy based on T_m-directed PBS length design and using a dual-pegRNA strategy remarkably increases PE efficiencies, and we developed a web application, PlantPegDesigner, based on these principles. This approach should greatly facilitate the application of PE in plants.

Online content

Any methods, additional references, Nature Research 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/s41587-021-00868-w.

Received: 22 August 2020; Accepted: 22 February 2021;
Published online: 25 March 2021

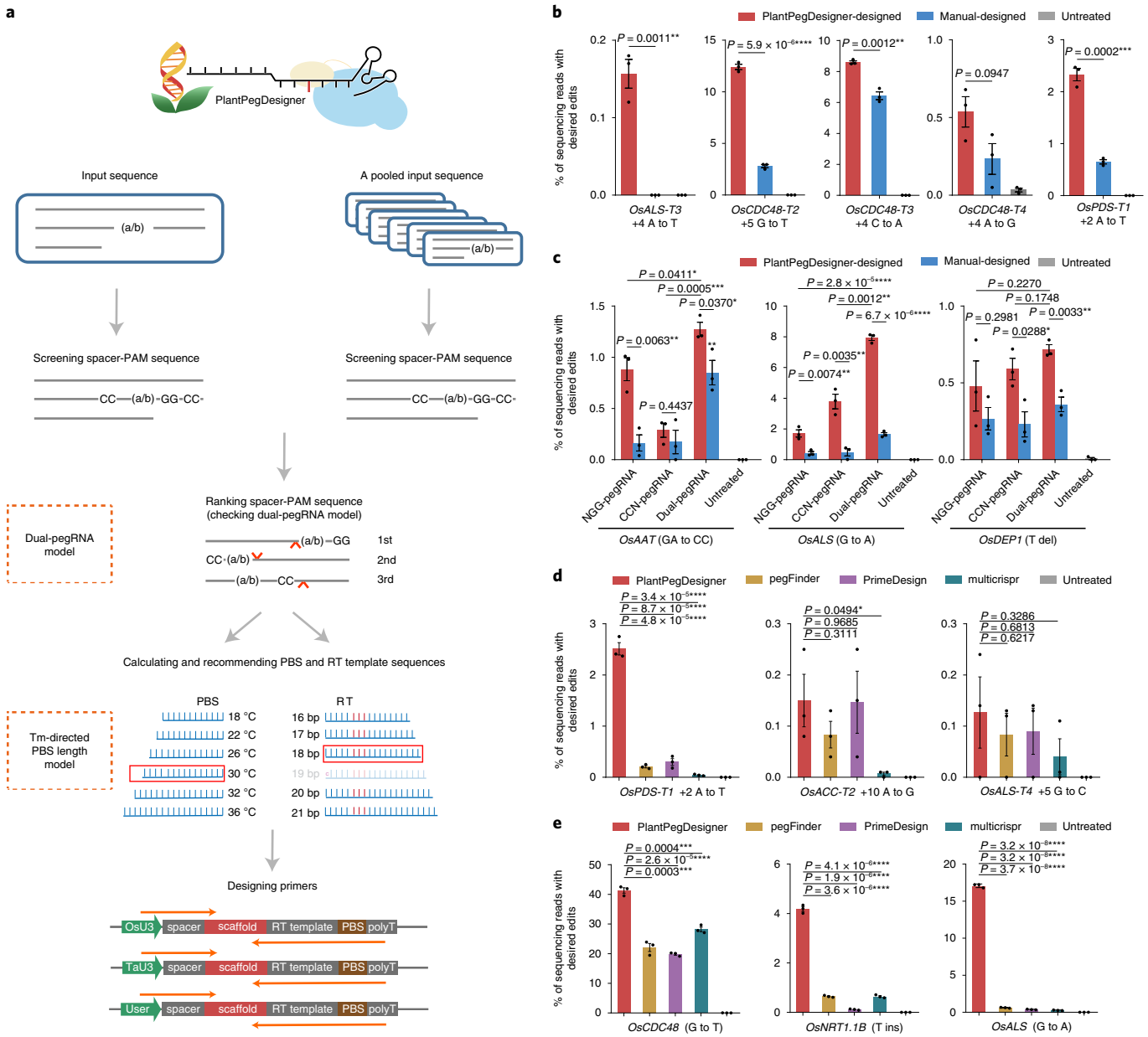

Fig. 2 | Workflow of the PlantPegDesigner web application and experimental validation. a, Schematic representation of the PlantPegDesigner workflow for designing pegRNAs. The user can provide a single input sequence or a file containing multiple input sequences to PlantPegDesigner. It screens and ranks the spacer-PAM sequences for installing the edits in the user-defined PE window, and the dual-pegRNA model is recommended if the spacer-PAM sequences can be found in both the forward and reverse strands for the input sequence. For each spacer-PAM sequence, PlantPegDesigner reports and recommends PBS sequences and RT template sequences based on PBS T_m and user-defined RT length, respectively. It also provides primers for pegRNA vector construction with appropriate plasmid vector scaffolds. **b**, **c**, Comparison of the editing efficiencies of PlantPegDesigner-designed and manual-designed (**b**) NGG-pegRNA or CCN-pegRNA and (**c**) dual-pegRNA in rice protoplasts. **d**, Comparison of the editing efficiencies of pegRNAs designed by PlantPegDesigner and three other web applications in rice protoplasts. **e**, Comparison of the editing efficiencies for the same edits using PlantPegDesigner-recommended dual-pegRNAs and three other web application-recommended single pegRNAs in rice protoplasts. The edits are referred to the base on the DNA forward strand in **c** and **e**. Frequencies (mean ± s.e.m.) in **b**–**e** are calculated from three independent experiments ($n = 3$). *P* values were obtained using two-tailed Student's *t*-tests. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$ and $^{****}P < 0.0001$.

References

1. Chen, K., Wang, Y., Zhang, R., Zhang, H. & Gao, C. CRISPR/Cas genome editing and precision plant breeding in agriculture. *Annu. Rev. Plant Biol.* **70**, 667–697 (2019).
2. Jiao, R. & Gao, C. Anything impossible with CRISPR/Cas9? *Sci. China Life Sci.* **60**, 445–446 (2017).
3. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
4. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat. Biotechnol.* **38**, 582–585 (2020).
5. Tang, X. et al. Plant prime editors enable precise gene editing in rice cells. *Mol. Plant* **13**, 667–670 (2020).
6. Li, H., Li, J., Chen, J., Yan, L. & Xia, L. Precise modifications of both exogenous and endogenous genes in rice by prime editing. *Mol. Plant* **13**, 671–674 (2020).
7. Xu, W. et al. Versatile nucleotides substitution in plant using an improved prime editing system. *Mol. Plant* **13**, 675–678 (2020).

8. Xu, R. et al. Development of plant prime-editing systems for precise genome editing. *Plant Commun.* **1**, 100043 (2020).
9. Hua, K., Jiang, Y., Tao, X. & Zhu, J. K. Precision genome engineering in rice using prime editing system. *Plant Biotechnol. J.* **18**, 2167–2169 (2020).
10. Butt, H. et al. Engineering herbicide resistance via prime editing in rice. *Plant Biotechnol. J.* **18**, 2370–2372 (2020).
11. Jiang, Y. et al. Prime editing efficiently generates W542L and S621I double mutations in two ALS genes of maize. *Genome Biol.* **21**, 257 (2020).
12. Liu, Y. et al. Efficient generation of mouse models with the prime editing system. *Cell Discov.* **6**, 27 (2020).
13. Surun, D. et al. Efficient generation and correction of mutations in human iPS cells utilizing mRNAs of CRISPR base editors and prime editors. *Genes* **11**, 511 (2020).
14. Kaback, D. B., Angerer, L. M. & Davidson, N. Improved methods for the formation and stabilization of R-loops. *Nucleic Acids Res.* **6**, 2499–2517 (1979).
15. Walton, R. T. et al. Unconstrained genome targeting with near-PAMless engineered CRISPR–Cas9 variants. *Science* **368**, 290–296 (2020).
16. Bhagwat, A. M. et al. multicrispr: gRNA design for prime editing and parallel targeting of thousands of targets. *Life Sci. Alliance* **3**, e202000757 (2020).
17. Chow, R. D., Chen, J. S., Shen, J. & Chen, S. A web tool for the design of prime-editing guide RNAs. *Nat. Biomed. Eng.* **5**, 190–194 (2020).
18. Hsu, J. Y. et al. PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat. Commun.* **12**, 1034 (2021).
19. Morris, J. A., Rahman, J. A., Guo, X. & Sanjana, N. E. Automated design of CRISPR prime editors for thousands of human pathogenic variants. Preprint at https://www.biorxiv.org/content/10.1101/2020.05.07.083444v1 (2020)
20. Kim, H. K. et al. Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.* **39**, 198–206 (2020).

Methods

Plasmid construction. The pegRNA constructs were made as reported previously[4]. NGG-pegRNA (NG-pegRNA) and CCN-pegRNA (CN-pegRNA) were cloned into the pOsU3 (digested with *BsaI* and *HindIII*) and pTaU3 vectors (digested with *Eps3I* and *NcoI*), respectively. The nSpG(H840A) was codon optimized for cereal plants and synthesized commercially (GENEWIZ), and the nSpG(H840A)-M-MLV fusion protein sequences were cloned into the vector pJIT163 backbone yielding the pSpG-PPE using a ClonExpress II One Step Cloning Kit (Vazyme). The pH-nCas9-PPE-V2 vector was modified from pH-nCas9-PPE[4] by blocking the *HindIII* site upstream of the *OsU3* promoter, which can be used for one-step pegRNA construction by digesting with *BsaI* and *HindIII*. PCR was performed using TransStart FastPfu DNA Polymerase (TransGen Biotech). All the primers were synthesized at the Beijing Genomics Institute.

Calculation of PBS T_m. The algorithm for computing the PBS T_m of pegRNAs was referred to the Oligo Analysis Tool (https://www.eurofinsgenomics.eu/en/ecom/tools/oligo-analysis/) when PBS length ≤ 15 nt. The formula used is

$$T_m = 4N_{G:C} + 2N_{A:T}$$

where $N_{G:C}$ and $N_{A:T}$ are the numbers of G:C and A:T base pairs in the PBS sequence, respectively.

Protoplast transfection. We used the *japonica* rice variety Zhonghua11 to prepare protoplasts. Protoplast isolation and transformation were performed as previously described[21,22]. Plasmids (10 μg per construct) were introduced by PEG-mediated transfection, with a mean transformation efficiency of 30–40%. Transfected protoplasts were incubated at 23 °C, and, after 48 h, they were collected and genomic DNA was extracted with the DNAquick Plant System (TIANGEN Biotech) used for deep amplicon sequencing (see below).

DNA extraction. Genomic DNA was extracted with a DNAquick Plant System. The targeted sequences were amplified with specific primers, and the amplicons were purified with an EasyPure PCR Purification Kit (TransGen Biotech) and quantified with a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific).

Amplicon deep sequencing and data analysis. Genomic DNA was extracted from protoplasts 48 h after transfection and used as template. In the first round of PCR, the target region was amplified from the protoplast DNA with site-specific primers (Supplementary Table 7). In a second round, both forward and reverse barcodes were added to the ends of the PCR products for library construction (Supplementary Table 7). Equal amounts of PCR products were pooled and sequenced commercially (Novogene) using the NovaSeq platform, and the pegRNA target sites in the sequenced reads were examined for substitutions and indels. Amplicon sequencing was repeated three times for each target site using DNA extracted from three independent protoplast samples. Analyses of PE processivity and byproducts were performed as previously described[4].

Development of the PlantPegDesigner algorithm and web server. The PlantPegDesigner algorithm was developed in Perl, and the web portal was implemented in PHP and JavaScript.

Statistical analysis. GraphPad Prism 8 software was used to analyze the data. All numerical values are presented as mean ± s.e.m. Differences between control and treatments were tested using two-tailed Student's *t*-tests. The efficiency in Fig. 1c was shown to follow a normal distribution according to the Kolmogorov–Smirnov test.

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

Data availability

All data supporting the findings of this study are available in the article and its supplementary figures and tables or can be obtained from the corresponding authors upon reasonable request. For sequence data, rice locus identifiers (http://rice.plantbiology.msu.edu/) are as follows: LOC_Os01g55540 (*OsAAT*), LOC_Os05g22940 (*OsACC*), LOC_Os03g54790 (*OsALS*), LOC_Os03g05730 (*OsCDC48*), LOC_Os09g26999 (*OsDEP1*), LOC_Os06g04280 (*OsEPSPS*), LOC_Os08g39890 (*OsIPA1*); LOC_Os10g40600 (*OsNRT1.1B*), LOC_Os08g03290 (*OsGAPDH*), LOC_Os03g08570 (*OsPDS*) and LOC_Os06g35970 (*OsROC5*). The deep sequencing data have been deposited in a National Center for Biotechnology Information BioProject database (accession code PRJNA702010). Plasmids pSpG-PPE, pOsU3, pTaU3 and pH-nCas9-PPE-V2 will be available through Addgene.

Code availability

The PlantPegDesigner web application code is available at GitHub (https://github.com/JinShuai001/PlantPegDesigner). The web portal server is accessible at http://www.plantgenomeediting.net for non-profit use.

References

21. Shan, Q. et al. Targeted genome modification of crop plants using a CRISPR–Cas system. *Nat. Biotechnol.* **31**, 686–688 (2013).
22. Zong, Y. et al. Efficient C-to-T base editing in plants using a fusion of nCas9 and human APOBEC3A. *Nat. Biotechnol.* **36**, 950–953 (2018).

Acknowledgements

This work was supported by grants from the National Transgenic Science and Technology Program of China (2019ZX08010-001 and 2019ZX08010-003), the Strategic Priority Research Program of the Chinese Academy of Sciences (Precision Seed Design and Breeding, XDA24020310 and XDA24030504) and the National Natural Science Foundation of China (31788103 and 32001060).

Author contributions

Q.L., S.J., Y.Z., J.Q., J.L. and C.G. designed the project. Q.L., Y.Z., S.J., Z.Z., G.L. and Y.W. performed the experiments. H.Y., S.J. and L.K. developed the web application. Q.L., S.J., Y.Z., J.Q., J.L. and C.G. wrote the manuscript. J.L. and C.G. supervised the project.

Competing interests

The authors have submitted a patent application based on the results reported in this paper.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-021-00868-w.

Correspondence and requests for materials should be addressed to J.L. or C.G.

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

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Caixia Gao

Last updated by author(s): Feb 16, 2021

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 Authors & Referees 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	Illumina NovaSeq platform was used to collect the amplicon deep sequencing data.
Data analysis	Graphpad prism 8 was used to analyze the data. Amplicon sequencing data of prime-editing processivity and indels was analyzed using the published code as previously described in reference 4. The custom Perl script to analyze types of mutational reads will be made available upon request. The URL of PlantPegDesigner is http://www.plantgenomeediting.net. The source code of PlantPegDesigner is deposited in Github at: https://github.com/JinShuai001/PlantPegDesigner.

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/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

The authors declare that all data supporting the findings of this study are available in the article and its Supplementary Information files or are available from the corresponding author on request. Datasets of high-throughput sequencing experiments have been deposited with the National Center for Biotechnology Information (NCBI) (accession code PRJNA702010).

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	The experiments of protoplasts were performed with three biological repeats. About 500,000 protoplasts were used for each transfection. The number of protoplasts in each transfection was measured by thrombocytometry.
Data exclusions	No data exclusion.
Replication	All attempts for replication were successful. For the experiments in rice protolasts, a minimum of three independent experiments were included.
Randomization	Rice protoplasts were isolated and randomly separated to each transformation.
Blinding	Not applicable. As samples were processed identically through standard and in some cases automated procedures (DNA sequencing, transfection, DNA isolation) that should not bias outcomes.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and*

	any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access and import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

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

Antibodies

Antibodies used	*Describe all antibodies used in the study; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Validation	*Describe the validation of each primary antibody for the species and application, noting any validation statements on the manufacturer's website, relevant citations, antibody profiles in online databases, or data provided in the manuscript.*

Palaeontology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).*

Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	*For laboratory animals, report species, strain, sex and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, gender, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	*Describe the antibodies used for the ChIP-seq experiments; as applicable, provide supplier name, catalog number, clone name, and lot number.*
Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

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	*Describe the sample preparation, detailing the biological source of the cells and any tissue processing steps used.*
Instrument	*Identify the instrument used for data collection, specifying make and model number.*
Software	*Describe the software used to collect and analyze the flow cytometry data. For custom code that has been deposited into a community repository, provide accession details.*
Cell population abundance	*Describe the abundance of the relevant cell populations within post-sort fractions, providing details on the purity of the samples and how it was determined.*
Gating strategy	*Describe the gating strategy used for all relevant experiments, specifying the preliminary FSC/SSC gates of the starting cell population, indicating where boundaries between "positive" and "negative" staining cell populations are defined.*

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

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s) — *Specify: functional, structural, diffusion, perfusion.*

Field strength — *Specify in Tesla*

Sequence & imaging parameters — *Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*

Area of acquisition — *State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software — *Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*

Normalization — *If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*

Normalization template — *Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*

Noise and artifact removal — *Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring — *Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.*

Statistical modeling & inference

Model type and settings — *Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).*

Effect(s) tested — *Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.*

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016) — *Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.*

Correction — *Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).*

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity — *Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).*

Graph analysis — *Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).*

Multivariate modeling and predictive analysis — *Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.*