A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing

Julian Grünewald, Ronghao Zhou, Caleb A. Lareau, Sara P. Garcia, Ronghao Zhou, Bret R. Miller, Lukas M. Langner, Jonathan Y. Hsu, Martin J. Aryee and J. Keith Joung

Existing adenine and cytosine base editors induce only a single type of modification, limiting the range of DNA alterations that can be created. Here we describe a CRISPR–Cas9-based synchronous programmable adenine and cytosine editor (SPACE) that can concurrently introduce A-to-G and C-to-T substitutions with minimal RNA off-target edits. SPACE expands the range of possible DNA sequence alterations, broadening the research applications of CRISPR base editors.

Adenine and cytosine base editors (ABEs and CBEs) enable programmable A-to-G and C-to-T transitions in DNA, respectively\(^1\)–\(^5\). To create a dual-function base editor, we engineered a single protein harboring adenosine and cytidine deaminases from the previously described miniABEmax-V82G (ref.\(^6\)) and Target-AID (ref.\(^5\)) editors, respectively (Fig. 1a and Extended Data Fig. 1a). We chose to combine parts of these particular editors because the deaminase domains were located on opposite ends of the \(\text{S}\text{pCas9}-\text{D}10\text{A}\) nickase in these fusions (Fig. 1a) and were previously shown to exhibit substantially reduced or minimal off-target RNA editing in human cells\(^6\). In addition, the monomeric TadA variant in miniABEmax-V82G is smaller than the dimeric wild-type TadA–TadA variant present in other ABEs, thereby helping to minimize the overall size of the resulting SPACE fusion. To maximize the purity of on-target cytosine base edits induced by SPACE, we also included two uracil glycosylase inhibitors (UGIs; Fig. 1a), a strategy first described with the previously published BE4 CBE (ref.\(^7\)).

We directly compared the on-target DNA editing activities of SPACE with those of miniABEmax-V82G and Target-AID using 28 different guide RNAs (gRNAs) in human HEK293T cells (Fig. 1b and Supplementary Table 1). Overall, SPACE induced A-to-G editing at 25 out of 26 genomic sites that were edited by miniABEmax-V82G, and C-to-T editing at all 28 genomic sites that were edited by Target-AID (Fig. 1b). The efficiencies of A-to-G editing by SPACE (range: 0.03–71.33%, mean: 13.07%) were modestly reduced relative to those of miniABEmax-V82G (range: 0.06–70.82%, mean: 18.1%), whereas the efficiencies of C-to-T editing by SPACE (range: 0.3–69.98%, mean: 22.13%) were comparable to those observed with Target-AID (range: 0.28–70.32%, mean: 24.82%). We did not observe a strict requirement by SPACE for the identity of the base present just 5’ to an edited A or C (Fig. 1b). The editing windows of SPACE were slightly narrowed compared to those of miniABEmax-V82G and Target-AID (Fig. 1c and Extended Data Fig. 1b,c). Among the 28 target sites we examined, A-to-G and C-to-T edits by SPACE were mostly induced at positions 4–7 and 2–7, respectively, within the protospacer (with 1 being the most distal base relative to the protospacer adjacent motif [PAM]; Fig. 1c). Our analysis also showed that for all 28 gRNAs, the product purities and indel frequencies at on-target sites edited by SPACE were comparable to or better than those of miniABEmax-V82G or Target-AID individually (Extended Data Figs. 2 and 3). Although we do not know the precise causes of modest differences in editing efficiencies we observed between SPACE and the individual miniABEmax-V82G or Target-AID editors, we note that there are differences in protein expression (codon and promoter usage), nuclear localization signal (NLS) architecture and the number of UGIs among these fusions (Extended Data Fig. 1a, Supplementary Table 2 and Methods).

An important potential advantage of SPACE relative to single-deaminase editors is the capability to concurrently introduce more than one type of base edit. Among the 25 gRNAs that induced both A-to-G and C-to-T on-target edits with SPACE, we observed that for 18 of these gRNAs the mean of the summed frequencies of dual-edited alleles was greater than 15% (Fig. 1d, Extended Data Fig. 4 and Supplementary Table 3). For 10 of these 25 gRNAs, the most efficiently edited on-target allele had both types of edit (for the replicate shown in Fig. 1d and Extended Data Fig. 4). To test whether SPACE is more efficient at inducing dual edits than the combined effects of separate ABEs and CBEs, we also performed experiments in HEK293T cells in which we directly compared SPACE with coexpressed miniABEmax-V82G and Target-AID (ABE & CBE mix) for each of the 28 gRNAs (Extended Data Fig. 5, Supplementary Table 4 and Methods). For 22 of the 28 gRNAs, the mean of the summed frequencies of dual-edited on-target alleles was higher with SPACE than with the ABE & CBE mix (Extended Data Fig. 6a and Supplementary Fig. 1). Interestingly, the mean of the summed frequencies of on-target alleles harboring only A-to-G edits was higher with SPACE than with the ABE & CBE mix conditioned with SPACE for 21 of these same 22 gRNAs (Extended Data Fig. 6a) whereas the mean of the summed frequencies of on-target alleles with only C-to-T edits was higher with SPACE than with the ABE & CBE mix for 16 of the 22 gRNAs (Extended Data Fig. 6a). The frequencies of unwanted indels induced by SPACE (range: 0.02–7.1%, mean: 1.44%) were lower or comparable to those observed with the ABE & CBE mix (range: 0.13–11.92%, mean: 2.88%) at 27 out of 28 sites tested (Extended Data Fig. 6b). Although we cannot rule out that differences in the architecture of SPACE, miniABEmax-V82G and Target-AID may affect their expression levels and/or activities (Supplementary Table 2), our results demonstrate that SPACE generally yields higher...
Fig. 1 | SPACE induces concurrent A-to-G and C-to-T base edits in human HEK293T cells. a, A schematic illustrating the ABE miniABEmax-V82G, the CBE Target-AID and the dual-deaminase base editor SPACE. A-to-G (pink halo) and C-to-T (blue halo) base edits are illustrated. The light blue shape indicates S. pyogenes Cas9 (D10A) nickase, the purple structure in Cas9 represents the gRNA, the pink circle represents the TadA 7.10 (V82G) adenosine deaminase monomer, the blue triangle indicates the Petromyzon marinus cytidine deaminase pmCDA1 and the yellow circles represent UGIs. The N and C labels denote the N and C termini of the base editor.

b, Heat maps showing the on-target DNA A-to-G (pink) and C-to-T (blue) editing frequencies of nCas9 (Control), miniABEmax-V82G, Target-AID and SPACE with 28 gRNAs (n = 4 independent replicates). The editing windows shown represent the most highly edited adenines and cytosines, not the entirety of the protospacer. The numbering at the bottom represents the position of the respective base in the protospacer sequence, with 1 being the most PAM-distal position.

c, Box and dot plots indicating the aggregate distribution of A-to-G (pink) and C-to-T (blue) edits across the entire protospacer with SPACE using 28 gRNAs. In the box plots, the box spans the interquartile range (IQR) (first to third quartiles), the horizontal line shows the median (second quartile) and the whiskers extend to ±1.5×IQR. The single dots represent individual replicates. The graph shows the same data as shown in b (n = 4).

d, The composition of alleles with frequencies of 1% or higher that result from SPACE-induced on-target DNA editing with a gRNA targeting HEK site 2 (ABE site 1). Data are shown for one replicate from the HEK site 2 on-target experiment shown in b. The numbering indicates the position in the protospacer, with 1 being the most PAM-distal position.
**Fig. 2 | RNA and DNA off-target editing, potential amino-acid modifications and number of genes with potential TF-binding sites induced by SPACE.**

a. Jitter plots showing transcriptomic A-to-I (pink) and C-to-U (blue) mutations detected in RNA-seq experiments from HEK293T cells in which ABEmax, miniABEmax-V82G, Target-AID or SPACE was coexpressed with either a HEK site 2 or RNF2 site 1 gRNA. The GFP control cells expressed no gRNA. Data are shown from three independent replicates. n represents the number of combined adenines and cytosines modified. b. Heat maps showing the A-to-G (pink) and C-to-T (blue) DNA off-target editing frequencies of nCas9 (Control), miniABEmax-V82G, Target-AID or SPACE coexpressed in HEK293T cells with gRNAs targeted to HEK sites 2-4, EMX1 site 1 or FANCF site 1 (n = 4 independent replicates). The respective off-target editing windows for each site are shown. The location in the protospacer is indicated at the bottom, with 1 being the most PAM-distal position. c. A Circos plot showing amino-acid changes that can be induced by SPACE (gray), including amino-acid changes in specific codon contexts, that are uniquely enabled (blue). d. Bar plots showing the computationally determined number of genes (y axis) that could be targeted by SPACE to install 1-10 (or more) TF-binding sites (x axis) in proximity of the TSS of coding genes in the human genome. Sites were filtered to contain a preferential SPACE editing window (C3-C4-A5, with 1 being the most PAM-distal position) and a canonical NGG PAM (Methods).
frequencies of dual-edited alleles and lower frequencies of indels at on-target sites compared to coexpression of standard editors harboring the same adenosine and cytidine deaminases individually.

To characterize the transcriptome-wide RNA off-target activity of SPACE, we performed RNA sequencing (RNA-seq) from HEK293T cells coexpressing SPACE with a gRNA targeting HEK site 2 or RNF2 site 1 (Methods). We also performed matched side-by-side RNA-seq experiments with HEK293T cells expressing miniABEmax-V82G, Target-AID, ABEmax (a positive control for RNA editing) and GFP (a negative control). Analysis of on-target DNA editing in cells from the same experiments showed efficient editing with SPACE, miniABEmax-V82G and Target-AID with both gRNAs (Extended Data Fig. 7a). As expected, GFP negative control experiments showed very few RNA C-to-U edits (range of 1–3) and A-to-I edits (range of 7–12) while ABEmax induced relatively high numbers of A-to-I edits (range of 3,105–5,696), miniABEmax-V82G induced low numbers of A-to-I edits (range of 73–194) and Target-AID induced even lower numbers of C-to-U edits (range of 6–11; Fig. 2a, Extended Data Fig. 7b and Supplementary Table 5). Cells expressing SPACE showed very few C-to-U edits (range of 0–4) and only small numbers of A-to-I edits (range of 4–37; Fig. 2a, Extended Data Fig. 7b and Supplementary Table 5). (The generally lower numbers of RNA edits that we observed in our current experiments relative to previously published studies39–42 are due to the reduced sequencing depth we used here (~14–18 million reads per sample) compared with our earlier work (~80–120 million reads per sample) (Methods).) On the basis of these results, we conclude that SPACE retains the reduced RNA-editing activities observed with miniABEmax-V82G and Target-AID, inducing very low numbers of unwanted RNA edits throughout the transcriptome.

We also assessed Cas9/gRNA-dependent DNA off-target activities of SPACE by using targeted amplicon sequencing to quantify editing frequencies at 23 previously defined Cas9 off-target sites for 5 gRNAs (targeting HEK sites 2, 3 and 4, EMX1 site 1 and FANC1 site 1)43. We directly compared SPACE editing frequencies at these 23 off-target sites to those observed with expression of miniABEmax-V82G, Target-AID or a Cas9-D10A nickase (as a negative control). For 17 of these 23 off-target sites, editing efficiencies observed with SPACE were comparable or lower relative to those observed with miniABEmax-V82G or Target-AID (Fig. 2b and Supplementary Table 6). This was also true for the remaining six off-target sites, with the exception of the adenines at positions 6 and 8 for HEK site 3 gRNA off-target site 1, the adenosines at position 4 for HEK site 4 gRNA off-target sites 1 and 4, the cytosine at position 6 for EMX1 site 1 gRNA off-target site 2, the cytosine at position 8 for FANC1 site 1 gRNA off-target site 1, and the adenine at position 4 for FANC1 site 1 gRNA off-target site 4 (Fig. 2b and Supplementary Table 6). We also confirmed that on-target DNA editing was present with the five different gRNAs and SPACE, miniABEmax-V82G or Target-AID coexpression (Extended Data Fig. 7c and Supplementary Table 6). This initial assessment suggests that SPACE does not induce dramatically different edit frequencies at known Cas9-dependent DNA off-target sites, but more comprehensive studies will be needed to fully define the genome-wide Cas9-dependent and gRNA-independent DNA off-target profiles of SPACE.

Our results demonstrate that SPACE enables the efficient and concurrent introduction of A-to-G and C-to-T edits, thereby expanding the range of targeted DNA edits that can be created with base editor technologies. This expanded editing capability will be broadly useful for a number of different research applications. For example, SPACE adds 60 additional codon changes (resulting in 18 amino-acid substitutions) that cannot be created with existing single-action CBes and ABEs (Extended Data Fig. 7d, Fig. 2c and Supplementary Table 7). In addition, SPACE could be useful for creating or reverting multi-nucleotide variants (MNVs), a newly emerging category of sequence variants associated with disease43,45 (also see ref. 11; Supplementary Table 8). Notably, among MNVs, TG-to-CA and CA-to-TG (both inducible by SPACE) are the most frequent consecutively arising adjacent dinucleotide MNVs. Furthermore, the greater combinatorial diversity of mutations that result with SPACE as compared with single-deaminase base editors could make it attractive for molecular recording systems (for example, lineage tracing)46,47 as well as for saturation mutagenesis screens, directed evolution and protein engineering48,49. Finally, we envision that the concurrent editing of cytosines and adenines may enable the programmable installation of specific CA/TG-rich transcription factor (TF)-binding sites. Our computational analysis indicates that thousands of genes harbor one or more SPACE-targetable CCA motifs in protospacer positions 3, 4 and 5 with respect to an NGG PAM, within a distance of ~500bp to 0bp relative to the transcription start site (TSS), that can be targeted to create TF-binding sites (Fig. 2d and Extended Data Fig. 8). Further protein engineering efforts may lead to expanded editing windows and higher editing efficiencies for future iterations of SPACE. While this work was being reviewed, another similar dual-deaminase architecture was reported by Gao and colleagues50. In summary, our development of SPACE should further expand the scope of research applications enabled by base editor technology.

Online content
Any 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/s41587-020-0535-y.

Received: 23 September 2019; Accepted: 23 April 2020; Published online: 1 June 2020

References
1. 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).
2. Gaudelli, N. M. et al. Programmable base editing of A to T in G/C genomic DNA without DNA cleavage. Nature 551, 464–471 (2017).
3. Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields CG-to-TA base editors with higher efficiency and product purity. Sci. Adv. 3, eaaq4774 (2017).
4. 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).
5. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, aaf7829 (2016).
6. Grünewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing activities. Nat. Biotechnol. 37, 1041–1048 (2019).
7. Grünewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. Nature 569, 433–437 (2019).
8. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. 33, 187–197 (2015).
9. Kaplanis, J. et al. Exome-wide assessment of the functional impact and pathogenicity of multinucleotide mutations. Genome Res. 29, 1047–1056 (2019).
10. Besenbacher, S. et al. Multi-nucleotide de novo mutations in humans. PLoS Genet. 12, e1006315 (2016).
11. Wang, Q. et al. Landscape of multi-nucleotide variants in 125,748 human exomes and 15,708 genomes. Preprint at bioRxiv https://doi.org/10.1101/573378 (2019).
12. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science 353, aaf7907 (2016).
13. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. Nature 570, 77–82 (2019).
14. Canver, M. C. et al. BCL11A enhancer dissection by Cas9-mediated in situ cleavage by CRISPR-Cas nucleases. Nature 569, 433–437 (2019).
15. Hoss, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat. Methods 13, 1036–1042 (2016).
16. Li, C. et al. Targeted, random mutagenesis of plant genes with dual cytidine and adenosine base editors. Nat. Biotechnol. https://doi.org/10.1038/s41587-019-0393-7 (2020).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020.
transfected 24 h post-seeding with 37.5
43.5
72 h post-transfection, cells were washed with PBS, and lysed in freshly prepared
For DNA on-target experiments in 96-well plates, at
DNA and RNA extraction.
µ
and filtered through 35-
0.05% SDS), 5.25
(0.05% SDS) for RNA extraction. Cells were centrifuged (200
×
g for RNA extraction. Cells were centrifuged (200
and gRNA plasmids
were cloned by ligation into the pUC19-based entry vector BPK1520 (Bsmbl
digest; Addgene no. 65777). All plasmids were mid or maxi prepped with the
Qagen Midi/Maxi Plus kits.
Cell culture. HEK293T cells (CRL-3216) were purchased from and authenticated using
cultured in Dulbecco’s modified Eagle medium ( Gibco) with 10 % (v/v) fetal bovine serum (FBS), 100 units/mL penicillin and 100
μg/mL streptomycin-pyrimethamine (Sigma) and passed every 3–4 days and tested for mycoplasma with MycoAlert PLUS (Lonza) every 4 weeks.
HEK293T cells were used until passage 20 for all experiments.
Transfections. For DNA on-target experiments with 28 gRNAs (Fig. 1b), 1.25 × 10⁴
HEK293T cells were seeded into 96-well flat-bottom cell culture plates (Corning),
transfected 24 h post-seeding with 150 ng base editor or control, 10 ng gRNA and
0.3 µl TransIT-X2 (Mirus) (per well), and collected 72 h after transfection to obtain
gDNA. For RNA off-target experiments, GFP sorted cells were split 20% for DNA and 80%
gRNA, and 1.5 µl TransIT-X2 (per well), and collected 72 h after transfection
by a NextSeq 500 machine using a NextSeq 500 High Output Kit v2 (150 cycles).
In summary, the first PCR amplified the genomic site of interest with primers
resuspended to 1
l DNA lysis buffer, 21
µl DNA lysis buffer (50 mM Tris HCl pH 8.0, 100 mM NaCl, 5 mM EDTA,
µl TransIT-293 (Mirus) (per dish), and sorted 36–40 h after transfection.
Annotated TSSs
NAC
TNG, GTN, ANC
TGN, CNA
NGT

conversions are defined as the reverse scenario, with targetable C and A bases
with matching T and G bases in the REF position; whereas disease-generating
disease-correcting or disease-generating modifications enabled by SPACE,
disease-generating or disease-generating modifications enabled by SPACE,
A-to-G/C-to-T edits include A-to-G/C-to-T edits identified on the positive strand
relative to
−
500 bp to 0 bp relative to the start site.
From this window of 500 bp per gene, we found all matches of the
NNCCANNNNNNNNNNNNGG motif on either strand that contain a
preferential dual-editing window for SPACE and a canonical SpCas9 PAM (NGG).
A-to-G/C-to-T edits include A-to-G/C-to-T edits identified on the positive strand.
A-to-G/C-to-T edits include A-to-G/C-to-T edits identified on the positive strand.

PCR added barcodes with primers containing unique pairs of barcodes that are analogous to Illumina TruSeq CD indexes. The libraries were pooled on the basis of Pico Green (Promega) measurements on a BioTek Synergy HT plate reader, and final pool was sequenced paired-end 2 x 150 on a MiSeq machine using a 300-cycle MiSeq Reagent Kit v2 (Illumina). Demultiplexed FASTQs were downloaded from Basespace (Illumina) and analyzed using the batch version of CRIPResso (ref. 1).
factor. Motif matching was performed using the motifmatchr package using default parameters as part of the chromVAR suite of tools\(^2\). Created motifs were those that did not occur in the reference sequence but were matches in the SPACE-edited sequence.

Circos plotting. Amino-acid and codon modification plots were constructed using Circos\(^8\).

Statistics and data reporting. Sample sizes were determined on the basis of the published work of others in the field who perform similar experiments and achieve reproducible results. The investigators were not blinded to experimental conditions or analysis of experimental data. We did not use any specific statistical tests. In the box plots, the box spans the IQR (first to third quartiles), the horizontal line shows the median (second quartile) and the whiskers extend to \(\pm 1.5 \times \text{IQR}\).

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

Data availability
Plasmids encoding SPACE have been deposited at Addgene (nos. 140242–140245). All RNA-seq next-generation sequencing data generated for this study have been deposited in the Gene Expression Omnibus data repository (series GSE137411). All targeted amplicon sequencing data (DNA on- and off-target editing) have been deposited at the Sequence Read Archive (PRJNA609075).

Code availability
The authors will make all previously unreported custom computer code used in this work available upon request.

References
17. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat. Biotechnol. 37, 224–226 (2019).
18. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
19. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).
20. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43, 491–498 (2011).
21. Karczewski, K.J. et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Preprint at bioRxiv https://doi.org/10.1101/531210 (2020) https://storage.googleapis.com/gnomad-public/release/2.1/mnv/gnomad_mnv_coding.tsv
22. Mathelier, A. et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 44, D110–D115 (2016).
23. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods 14, 975–978 (2017).
24. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009).

Acknowledgements
Support for this work was provided by the National Institutes of Health (RM1 HG009490 to J.K.J. and R35 GM118158 to J.K.J and M.J.A.). J.K.J. is additionally supported by the Desmond and Ann Heathwood MGH Research Scholar Award and the Robert B. Colvin, M. D. Endowed Chair in Pathology. J.G. was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) – Projektnummer 416375182. L.M.L. was supported by a Boehringer Ingelheim Fonds MD fellowship. We thank M. K. Clement for technical advice, K. Petri and P. K. Cabeceiras for discussions and technical advice, and I. Paul-Pottenplackel for assistance with editing the manuscript.

Author contributions
C.A.L., S.P.G. and S.I. contributed equally to this work and are co-second authors. Wet laboratory experiments were performed by R.Z., B.R.M., L.M.L. and J.G. C.A.L., S.P.G., S.I., J.Y.H. and M.J.A. performed computational analyses. J.G., R.Z. and J.K.J. conceived of and designed the study. J.G., M.J.A. and J.K.J. supervised the work. J.G., R.Z. and J.K.J. wrote the initial manuscript draft and all authors contributed to the writing of the final manuscript.

Competing interests
J.K.J. has financial interests in Beam Therapeutics, Editas Medicine, Excelsior Genomics, Pairwise Plants, Poseida Therapeutics, Transposagen Biopharmaceuticals and Verve Therapeutics (f/k/a Endcadia). M.J.A. has financial interests in Excelsior Genomics. The interests of J.K.J. and M.J.A. were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. J.K.J. is a member of the Board of Directors of the American Society of Gene and Cell Therapy. J.G., R.Z. and J.K.J. are co-inventors on a patent application that has been filed by Partners Healthcare/Massachusetts General Hospital in accordance with their conflict of interest policies. J.K.J. is a member of the Board of Directors of the American Society of Gene and Cell Therapy. J.G., R.Z. and J.K.J. are co-inventors on a patent application that has been filed by Partners Healthcare/Massachusetts General Hospital in accordance with their conflict of interest policies.
Extended Data Fig. 1 | Architectures of miniABEmax-V82G, Target-AID, and SPACE, and A-to-G or C-to-T editing distributions of miniABEmax-V82G or Target-AID. a, Schematic illustration of miniABEmax-V82G, Target-AID, and SPACE architectures. Orange boxes = bipartite NLS for miniABEmax-V82G and SPACE or NLS for Target-AID; TadA* = mutant TadA 7.10 with V82G mutation, light grey box = SH3-3xFLAG for Target-AID, CDA1 = pmCDA1 with R187W mutation, and yellow boxes = UGis. b, c, Box and dot plots indicating the distributions of A-to-G (pink, b) and C-to-T (blue, c) edits across 28 pooled genomic sites with miniABEmax-V82G (b) and Target-AID (c) including the entire protospacer. In box plots, the box spans the interquartile range (IQR) (first to third quartiles), the horizontal line shows the median (second quartile), and the whiskers extend to ±(1.5 x IQR). Single dots represent individual replicates. Graph was made using the same data as shown in Fig. 1b (n = 4).
**Extended Data Fig. 2** | On-target C-to-R (A/G) and A-to-Y (C/T) editing of miniABEmax-V82G, Target-AID, and SPACE. Box and dot plots showing on-target DNA C-to-R (A/G) and A-to-Y (C/T) editing frequencies of miniABEmax-V82G (pink), Target-AID (blue), and SPACE (green) with 28 gRNAs (n=4). Box, horizontal line, and whiskers are defined as in Fig. 1c. Single dots represent all Cs or As across the entire protospacer for all four replicates at each genomic site. Data from the same experiment as shown in Fig. 1b. Please note that box plots are mostly contained within the horizontal line (median, second quartile) close to or at the value ‘0’.
Extended Data Fig. 3 | On-target indel frequencies induced by miniABEmax-V82G, Target-AID, and SPACE. Dot plots showing on-target DNA indel frequencies induced by nCas9-D10A (Control, black), miniABEmax-V82G (pink), Target-AID (blue), and SPACE (green) with 28 gRNAs (n = 4). Single dots represent individual replicates. Data from the same experiment as shown in Fig. 1b.
Extended Data Fig. 4 | Allele frequency tables of DNA on-target editing by SPACe. Composition of alleles with frequencies of 1% or higher that result from SPACe editing with 27 gRNAs (HEK site 2 data shown in Fig. 1d). Data are taken from the first replicate obtained for each gRNA from the on-target experiment shown in Fig. 1b. Numbering indicates the position in the protospacer with 1 being the most PAM-distal location.
Extended Data Fig. 5 | On-target DNA editing of SPACE compared to coexpression of miniABEmax-V82G and Target-AID (ABE & CBE mix) at 28 genomic sites. Heat maps showing on-target DNA A-to-G (pink) and C-to-T (blue) editing frequencies induced by nCas9 (Control), coexpression of miniABEmax-V82G and Target-AID (ABE & CBE mix) or SPACE with 28 gRNAs (n = 4 independent replicates). Editing windows shown represent the edited adenines and cytosines, not the entirety of the protospacer. Numbering at the bottom represents the position of the respective base in the protospacer sequence with 1 being the most PAM-distal location.
Extended Data Fig. 6 | On-target C-to-T, A-to-G, and dual editing, and indel frequencies induced by coexpression of miniABEmax-V82G and Target-AID compared with SPACE. a, Bar and dot plots showing mean sum of allele frequencies of all edited alleles with A-to-G only, C-to-T only, and concurrent A-to-G and C-to-T editing resulting from coexpression of miniABEmax-V82G and Target-AID (ABE & CBE mix, grey) or SPACE (green) with 28 gRNAs (n = 4). Single dots represent individual replicates. Error bars represent the standard deviation (SD). Data are from the same experiment as shown in Extended Data Fig. 5. b, Dot plots showing on-target DNA indel frequencies of nCas9 (Control, black), coexpression of miniABEmax-V82G and Target-AID (ABE & CBE mix, grey) and SPACE (green) with 28 gRNAs (n = 4). Single dots represent individual replicates. Data are from the same experiment as shown in Extended Data Fig. 5.
Extended Data Fig. 7 | Additional data and analysis from DNA and RNA off-target experiments and SPACE-inducible codon and amino-acid modifications. a, Heat maps showing the on-target DNA A-to-G (pink) and C-to-T (blue) editing frequencies of nCas9 (Control), ABEmax, miniABEmax-V82G, Target-AID, or SPACE with HEK site 2 and RNF2 site 1 gRNAs (n = 3 independent replicates) for the RNA-seq experiments shown in Fig. 2a. Editing windows shown represent the most highly edited adenines and cytosines, not the entire protospacer. Numbering at the bottom represents the position of the respective base in the protospacer sequence with 1 being the most PAM-distal location. b, Histograms showing the total number of RNA A-to-I or C-to-U edits observed (y-axis) with different editing efficiencies (x-axis) for ABEmax, miniABEmax-V82G, Target-AID, or SPACE, each tested with the HEK site 2 and RNF2 site 1 gRNAs. n = number of modified adenines and cytosines. Experiments were performed in triplicate (data are from the same experiments as shown in Fig. 2a). Dashed red line, median; solid red line, mean. c, Heat maps showing on-target DNA A-to-G (pink) and C-to-T (blue) editing efficiencies of nCas9 (Control), miniABEmax-V82G, Target-AID, or SPACE with HEK sites 2–4, EMX1 site 1, and FANCF site 1 gRNAs (n = 4 independent replicates) for DNA off-target experiments shown in Fig. 2b. Editing windows shown represent the most highly edited adenines and cytosines, not the entire protospacer. Numbering at the bottom represents the position of the respective base in the protospacer sequence with 1 being the most PAM-distal location. d, Circos plot showing 60 unique codon changes (with respect to the start codon) that can be induced by dual editing of adenines and cytosines by SPACE (grey), 18 of which (blue) lead to unique SPACE-inducible amino-acid changes with respect to the original codon (also see Fig. 2c and Supplementary Table 7).
Extended Data Fig. 8 | Potential transcription factor binding sites that can be created with SPACE. Bar plot showing computationally determined number of genes (y-axis) that could be targeted by SPACE to install 1–5 (or more) transcription factor binding sites (x-axis) in proximity of the transcription start site of coding genes in the human genome for 10 transcription factors. Sites were filtered to contain a preferential SPACE editing window (C3-C4-A5) and a canonical NGG PAM (Methods).
Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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 |
|   | An indication of 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 statistics 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 |
|   | Clearly defined error bars |
|   | State explicitly what error bars represent (e.g. SD, SE, CI) |

Our web collection on statistics for biologists may be useful.

Software and code

|   |   |
|---|---|
| Data collection | Next-generation sequencing data was collected with Illumina MiSeq (targeted DNA amplicon sequencing) and NextSeq (RNA-seq) instruments. FACS data was generated using a BD FACSARia II. For fluorometric assays (Pico Green) we used the Synergy HT microplate reader (BioTek) and the GenS software. |
| Data analysis | BD FACSDiva Software version 6.1.3, Microsoft Excel Version 1808 (Build 10730.20304), CRISPResso2 (2.0.30), bam-readcount version 0.8.0, STAR 2.6.0c, GATK 3.8, bowtie2 version 2.3.1, Picard version 2.7.1, 1000genomes version phase3, ESP version V2-SSA137, gnomAD version 170228, GENCODE version 28, genebuild version 2014-07, HGMD-PUBLIC version 20174, regbuild version 20190821 and dbSNP version 150, MSigDB database version 6.2; GSEA/MSigDB website version 6.3, Python version 2.7, R version 3.5, GenS 1.1.15, Circos 0.69-9, R packages chromVAR (v.1.8.0), motifmatchr (v1.8.0), JASPAR2016 (v1.14.0), Biostrings version2.50.2, data.table version 1.12.2, dplyr version 0.8.3, ggplot2 version 2.17.0, ggrepel version 3.2.1, RColorBrewer version 1.1-2, stringr version 1.4.0, tidyverse version 1.2.1., Illumina BaseSpace |

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 upon request. 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 RNA-seq data generated for this study have been deposited in the GEO data repository (series GSE137411), accessible via: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137411. All targeted amplicon sequencing data (DNA on- and off-target editing) have been deposited under the BioProject accession number PRJNA609075 at the Sequence Read Archive (SRA), accessible via: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA609075.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size
Sample sizes were determined based on other groups' work who generate reproducible results in similar experimental contexts.

Data exclusions
No data were excluded.

Replication
Independent replicates were performed with different cell passages/batches.

Randomization
Samples were not randomized. Controls were run in parallel whenever applicable to control for covariates.

Blinding
Blinding was not performed.

Reporting for specific materials, systems and methods

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☑   | Unique biological materials |
| ☑   | Antibodies |
| ☑   | Eukaryotic cell lines |
| ☑   | Palaeontology |
| ☑   | Animals and other organisms |
| ☑   | Human research participants |

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)
source: ATCC; cell lines used: HEK293T cells (ATCC CRL-3216)

Authentication
STR profiling by ATCC (December 2018)

Mycoplasma contamination
Supernatant was analyzed every 4 weeks using MycoAlert PLUS (Lonza). Cells continuously tested negative.

Commonly misidentified lines (See ICLAC register)
HEK293T are not listed in the ICLAC register (version 10).
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**
Cell culture, transfection and sorting are described in the online methods. 36 hours after transfection, cells were washed, trypsinized, and filtered through a 35μm cell strainer cap before FACS.

**Instrument**
FACS Aria II (BD Biosciences)

**Software**
BD FACSDiva Software v6.1.3

**Cell population abundance**
Post-gating cell population abundances for target populations were similar across different base editor experiments. HEK293T cells transfected with CBE, ABE, or dual-deaminase base editor plasmids co-translationally expressing EGFP via P2A (please see FACS figures in Supplementary Information) usually were ~40-50% GFP+ (% parent in BD FACSDiva), while cells expressing only EGFP (no base editor) were ~70-80% GFP-positive.

**Gating strategy**
Gates were established using untransfected control cells and transfected GFP+ cells. Gates were drawn to collect all GFP+ cells for base editor and control conditions.

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