nature research

Corresponding author(s): Dali Li

Last updated by author(s): Mar 21, 2020

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	Targeted amplicon sequencing data data was collected and demultiplexed by an Illumina HiSeq X Ten instrument. RNA-seq data was collected and demultiplexed by an Illumina NovaSeq 6000 instrument. FACS gating data was collected on a FACSAria III (BD Biosciences) using FACSDiva version 8.0.2 (BD Biosciences)
Data analysis	High-throughput sequencing data was analyzed by BE-Analyzer (http://www.rgenome.net/be-analyzer/#!)(Hwang G-H et al, BMC Bioinformatics, 2018) or CRISPResso2 (http://crispresso.pinellolab.partners.org/)(Clement, K. et al.Nat Biotechnol. 2019)for base editing(C>T or A>G or C>T and A>G) and indels efficiencies. Potential DNA off-targets site for A&C-BEmax were predicated using cas-OFFinder web (http://www.rgenome.net/cas-offinder/). RNA-seq data were analyzed using in-house Perl scripts, Hisat2 v2.0.5 and GATK (v4.0) software . FACS data was analyzed using FlowJo v.10. Prism 7 was also used to analyze data.

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

Targeted amplicon sequencing data have been deposited in the NCBI Sequence Read Archive database under Accession Code PRJNA558944, PRJNA559260, PRJNA559237, PRJNA559051, PRJNA592341 and PRJNA592333. The RNA-seq data used in this study have been deposited in the NCBI Sequence

Read Archive database under Accession Code PRJNA592597.

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 statistical methods were used to predetermine sample size.Experiments were performed in biological triplicate n=3 unless otherwise noted.
Data exclusions	No data was excluded.
Replication	Three independent biological replicates were performed on different days. All replications were successful.
Randomization	Samples were randomly distributed into groups.
Blinding	One person collected samples. Another person analyzed without any description.

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	a 1:10000 dilution of Anti-GAPDH (Abcam, ab9485) a 1:5000 dilution of anti-CRISPR-Cas9 (Abcam, ab189380).
Validation	The GAPDH antibody has been validated by western blot in HEK293T cells lysate(https://www.abcam.com/gapdh-antibody-loading-control-ab9485.html). The Cas9 (ab204448, Abcam) antibody has been validated by western blot in HEK293T cells transfected with CRISPR-Cas9 (Q99ZW2, Streptococcus pyogenes serotype M1) with GFP-Myc tag (https://www.abcam.com/crispr-cas9-antibody-epr18991-ab189380.html).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK 293T (purchased from ATCC CRL-3216) Hela (purchased from ATCC® CCL-2™) HUDEP-2 cell line (Ryo Kurita and Yukio Nakamura, Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki,Japan) HUDEP-2(ΔGγ) cell line (University of New South Wales, Merlin Crossley lab)
Authentication	HEK293T cells were directly purchased from ATCC with certification; HUDEP-2(ΔGγ) cells were confirmed by PCR and differentiation studies followed by FACS analysis.

Mycoplasma contamination	All cell lines used were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were 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	Cell culture and transfection procedures are described in the online methods. Cells were washed and passed through a 45µm cell strainer cap before sorting (72h after transfection).
Instrument	FACSAria III (BD Biosciences)
Software	BD FACSDiva Software Diva8.0.2
Cell population abundance	HEK293T Cell population abundances after gating for target populations were similar in different biology replicates. HEK293T cells infected with base editors described in the supplement usually were ~ 50-60% GFP+ (of gated population = % parent in BD FACSDiva).
Gating strategy	Gates were established using uninfected control cells and GFP positive control. Gates were drawn to collect subsets of GFP-expressing cells. cells with top 8% of GFP signal were sorted, after gating for the cell population (~8% of parent). Please see the Supplementary Information for gating strategies in different biology replicates.

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