nature research

Corresponding author(s): David R. Liu

Last updated by author(s): Oct 8, 2019

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 Miseq Control software (3.1) was used on the Illumina Miseq sequencers to collect the high-throughput sequencing data
Data analysis	Crispresso2 was used to analyze HTS data for quantifying editing activity at genomic sites. Cell Sorter Software Version 3.0.5 was used for flow cytometry analysis. RNA-seq demultiplexing was performed with bcl2fastq2 version 2.20, and sequences were trimmed with TrimGalore v. 0.6.2. Alignment of RNA-seq reads to the human genome was performed with RSEM version 1.3.1. RNA-seq data output was genearted with limma-voom and visualized in R. Frequency, mean, and standard deviations were calculated using GraphPad Prism 8. Custom python scripts provided in Supplementary Note 4 were used to analyze and quantify guide RNA scaffold insertion.

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

High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database under accession code PRJNA565979.

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	Sample sizes were determined based on literature precedence for genome editing experiments.
Data exclusions	No data was excluded.
Replication	All experiments were repeated at least once. All attempts at replication were successful.
Randomization	Yeast and mammalian cells used in this study were grown under identical conditions; no randomization was used.
Blinding	Yeast and mammalian cells used in this study were grown under identical conditions; blinding was not used.

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (ATCC), U2OS (ATCC), K562 (ATCC), HeLa (ATCC).
Authentication	Cells were authenticated by the supplier using STR analysis.
Mycoplasma contamination	All cell lines tested negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	None used.

Animals and other organisms

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

Laboratory animals	To generate dissociated neuronal cultures, timed-pregnant C57BL/6 mice were provided by Charles River. Pregnant mice were euthanized at E18.5, and tissue for dissociated cultures was harvested from all embryos.
Wild animals	The study did not involve wild animals.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	The Broad IACUC provided ethical guidance.

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

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	2.5 x 104HEK293T cells grown in the absence of antibiotic were seeded on 48-well poly-D-lysine coated plates (Corning). 16-24 h post-seeding, cells were transfected at approximately 70% confluency with 1 µL of Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's protocols and 750 ng of PE2-P2A-GFP plasmid, 250 ng of pegRNA plasmid, and 83 ng of sgRNA plasmid. After 3 days post transfection, cells were washed with phosphate-buffered saline (Gibco) and dissociated using TrypLE Express (Gibco). Cells were then diluted with DMEM plus GlutaMax (Thermo Fisher Scientific) supplemented with 10% (v/v) FBS (Gibco) and passed through a 35-µm cell strainer (Corning) prior to sorting. Cells were treated with 3 nM DAPI (BioLegend) 15 minutes prior to sorting.
Instrument	Sony LE-MA900 Cell Sorter
Software	Cell Sorter Software Version 3.0.5 (Sony)
Cell population abundance	Of the surviving single sorted HEK293T cells edited to have HEXA 1278+TATC, 3.02% were homozygous. Of the surviving single sorted HEK293T cells edited to have HBB E6V, 25% were homozygous. Cells were genotyped using next-generation sequencing (Illumina).
Gating strategy	HEK293T cells were initially gated on population using FSC-A/BSC-A (Gate A) and then sorted for singlets using FSC-A/FSC-H (Gate B). Live cells were sorted for by gating for DAPI-negative cells (Gate C). Finally the upper 50% of GFP expressing cells were sorted for using eGFP as the fluorochrome (Gate D).

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