nature research

Corresponding author(s): David R. Liu

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.

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

Policy information about availability of computer code

Data collection	A custom python script, provided in the SI, was used to assess SNPs targetable by base editors on the ClinVar database.
Data analysis	The burrows-wheeler aligner (bwa-mem) and pysamstats are two publically available data analysis tools for HTS analysis. Ancestral reconstruction used: IQ-TREE v 1.6.1, PAML – v 1.8 , and MAFFT – Ubuntu version 7.394. 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 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

Plasmids encoding BE4max, AncBE4max, and ABEmax have been deposited to Addgene. High-throughput sequencing data are deposited in the NCBI Sequence Read

Archive (SRP145378). Accession numbers for rat APOBEC sequences are provided in supplementary file 1 and supplementary sequences 1. There are no restrictions on data availability.

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	Experiments were performed in biological triplicate n=3 unless otherwise noted. In previous studies using related experiments we determined this sample size to be sufficient to ensure reproducibility. No statistical tests were used to determine sample size.
Data exclusions	No data was excluded.
Replication	All attempts at replication were successful, and standard deviations were within expected ranges.
Randomization	Different cell passages were used for each biological replicate.
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.

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

Antibodies

Antibodies used	Rabbit anti-HA antibody from Cell Signaling Technology, product # 3724S, Lot 8, C2954 used at 1:1000 dilution; Rabbit anti-B-actin from Cell Signaling, product # 4970S, Lot 14, 13E5 used at 1:200 dilution.
Validation	Validation was performed by supplier.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T (ATCC), N2A (ATCC), Fibroblast (Coriell, GM20958)
Authentication	Cells were authenticated by the supplier.
Mycoplasma contamination	HEK293T cells tested negative for mycoplasma.
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	Cells were grown in tissue culture, transfected or nucleofected as described in the text and grown for 3 days. Cells were then trypsinized and filtered to remove debris before sorting.
Instrument	FACSAria II
Software	BD FACS DIVA software was used for analysis.
Cell population abundance	The abundance of cells in the post-sort fraction was dependent upon cell type and sorting condition. HEK293T mCherry+ in trans cells were typically over 50% of the population, GFP/mCherry dual+ cells were also typically over 50% of the population, BE4-P2A-GFP GFP/mCherry dual+ cells were typically over 20% of the population, BE4max-P2A-GFP and ancBE4max-P2A-GFP GFP/mCherry dual+ cells were typically over 40% of the population. N2A In trans mCherry+ cells were typically over 60% of the population. N2A GFP+ cells for BE4max-P2A-GFP and ancBE4-P2A-GFP were typically over 30% of the population. Fibroblast in trans mCherry+ cells were typically over 20% of the population. Fibroblast GFP+ cells for BE4max-P2A-GFP and ancBE4max-P2A-GFP were typically over 10% of the population.
Gating strategy	Negative control (unstained) and fluorophore-positive cells were used to establish gates for each cell type. Gates were drawn to collect cells expressing either fluorophore. See the provided examples for gates used.

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