

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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☒ | ☐ 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
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ 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

EditR 1.0.8, FlowJo X, FastQC (v0.11.3), Trimmomatic (v0.36), Illumina HiSeq X-Ten

Data analysis

Microsoft Excel 2019, GraphPad Prism 8.0.2, STAR (v2.5.2b), Picard tools (v2.3.0), SplitNCigarReads, IndelRealigner, BaseRecalibrator and HaplotypeCaller tools from GATK (v3.5), Variant Effect Predictor (VEP, v94), RSEM (v1.2.21)

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

All raw data are deposited in NCBI Sequence Read Archive (SRA) under project accession PRJNA528149 and PRJNA528561 or <http://www.biosino.org/node/project/detail/OEP000277> and interpreted data is either available in the supplemental information or upon request.

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](https://www.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 size was chosen based on the results of previous works in this field that used similar sample sizes to generate reproducible results.
Data exclusions	To avoid mistakes during sample preparation and sequencing, samples expressing the right transfected vectors were included in this study. One RNA-seq dataset was excluded due to expression of the wrong transfected vectors.
Replication	To ensure the robustness, we performed independent biological replicates for each experiments. Moreover, we replicated the results using different gRNAs.
Randomization	Randomization was not relevant to this study given that top 5% GFP positive cells were collected by FACS sorting strategy for all samples.
Blinding	Blinding was not relevant to our study. Based on previous studies in this field, these assays do not require blinding. Thus, 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

Methods

n/a	Involved in the study	n/a	Involved in the study
☒	☐ Antibodies	☒	☐ ChIP-seq
☐	☒ Eukaryotic cell lines	☐	☒ Flow cytometry
☒	☐ Palaeontology	☒	☐ MRI-based neuroimaging
☒	☐ Animals and other organisms		
☒	☐ Human research participants		
☒	☐ Clinical data		

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	In this study, we only used one cell line HEK293T, which was from the Cell bank of Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences
Authentication	Cell lines were authenticated by the supplier.
Mycoplasma contamination	Cells used in this study were free of mycoplasma contamination. Mycoplasma contamination was determined by PCR the supernatant of HEK293T cells.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used was listed in the database of ICLAC.

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	HEK293T cells were digested by trypsin (0.05%), centrifuged at 1000 rpm and filtered with a 35 µm nylon mesh.
Instrument	MoFlo XDP (Beckman)
Software	Summit Software version 5.2, FlowJo X.
Cell population abundance	Cell population abundance: Cell population abundance was influenced by the size of the plasmids. Normally, HEK293T cells transfected with plasmids were usually ~60% GFP+ and around 500000 cells were sorted for RNA-seq.
Gating strategy	Positive and negative boundaries were determined by control cells that were not transfected with any plasmids. Cells with top 5% of GFP signal were collected.

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