nature research

Corresponding author(s): J. Keith Joung

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	Next-generation sequencing data was collected and demultiplexed by Illumina HiSeq2500, NextSeq500 (RNA-seq), NovaSeq (WES) and MiSeq (targeted amplicon sequencing) instruments. FACS data was generated using a BD FACSAria II. For fluorometric assays (Pico Green) and luciferase assays (CellTiter-Glo), we used the Synergy HT microplate reader (BioTek) using Gen5 software.
Data analysis	BD FACSDiva Software version 6.1.3, Microsoft Excel Version 1808 (Build 10730.20304), CRISPResso2 (release 20180918), bam-readcount version 0.8.0, STAR 2.6.0c, GATK 3.8, bowtie2 version 2.3.1, Picard version 2.7.1, Ensembl Variant Effect Predictor version 92.5, Polyphen version 2.2.2, Sift version 5.2.2, COSMIC version 83, 1000genomes version phase3, ESP version V2-SSA137, gnomAD version 170228, GENCODE version 28, genebuild version 2014-07, HGMD-PUBLIC version 20174, regbuild version 16, ClinVar version 201802, and dbSNP version 150, WebLogo 2.8 and 3.6.0, MSigDB database version 6.2; GSEA/MSigDB website version 6.3, biomaRt version 2.38.0, netMHC version 4.0, Python version 2.7, R version 3.5, GraphPad Prism 8 (part of cell viability data analysis and plotting), Variant Effect Predictor (VEP) version 92.5, Gen5 1.11.5.

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 next-generation sequencing data has already been deposited at the GEO data repository at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE121668. All WES and targeted amplicon sequencing data have been deposited at the SRA repository at https://www.ncbi.nlm.nih.gov/bioproject/PRJNA497753.

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	We determined sample sizes based on the results of other groups in the field who have used similar or identical sample sizes with these assays to generate reproducible results.
Data exclusions	We excluded one RNA-seq dataset due to sample impurity (negative control and experiment mixed during FACS).
Replication	We tested experimental conditions in different cell lines, using different gRNAs, and using different sequencing technologies to ensure robustness. We also performed biological replicates performed independently of each other.
Randomization	The samples were not randomized. Different cell passages/batches were used for distinct biological replicates. Covariates were controlled for by running controls in parallel whenever applicable.
Blinding	Blinding was not relevant to our study because in general, based on the prior experience of other groups in the field, these types of assays do not require blinding.

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) and HepG2 cells (ATCC HB-8065)
Authentication	STR profiling by ATCC
Mycoplasma contamination	Supernatant was analyzed bi-weekly using MycoAlert PLUS (Lonza). Both cell lines continuously tested negative. Additionally, we used RNA-seq data analysis to rule out mycoplasma infection a posteriori.

Commonly misidentified lines (See ICLAC register)	HEK293T and HepG2 cell lines are not listed in the ICLAC register (version 9, downloaded 27th November 2018).

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 and transfected as described in the online methods. 36-40h after transfection cells were washed, trypsinized and filtered through a 35µm cell strainer cap before starting the sorting procedure.
Instrument	FACSAria II (BD Biosciences)
Software	BD FACSDiva Software v6.1.3
Cell population abundance	Cell population abundance after gating for target populations depended on cell type, transfection and sorting conditions. HEK293T cells transfected with pCAG-nCas9-UGI-NLS-P2A-EGFP or pCAG-BE3-P2A-EGFP (including variations thereof) were usually ~ 40-60% GFP+ (of gated population = % parent in BD FACSDiva). HepG2 cells were usually 15-25% GFP+ after transfection with a pCAG-P2A-EGFP control plasmid. Usage of larger plasmids like pCAG-nCas9-UGI-NLS-P2A-EGFP or pCAG-BE3-P2A-EGFP (including variations thereof) led to reduced transfection efficiency, reflected in GFP+ populations of ~10-20%. In both cell lines, BE3-based plasmids showed slightly lower GFP+ cells and/or peak GFP signal than nCas9-UGI-NLS-based controls due to larger plasmid size and thus reduced transfection efficiency. For ABEmax experiments, HEK293T cells transfected with pCMV-bpNLS-32AA linker-nCas9(D10A)-bpNLS-P2A-EGFP (negative control) or pCMV-ABEmax-P2A-EGFP were usually ~ 40-60% GFP+ (of gated population = % parent in BD FACSDiva).
Gating strategy	We used untransfected control cells and transfected GFP+ cells to establish gates for both HEK293T and HepG2 cells. We drew gates to collect either all GFP+ cells or subsets of GFP-expressing cells. When sorting for highly GFP-expressing cells, we collected cells with top 5% of GFP signal of all cells after initial gating for the cell population (~5% of parent). For nCas9-UGI-NLS-P2A-EGFP, bpNLS-32AA linker-nCas9(D10A)-bpNLS-P2A-EGFP controls or P2A-EGFP controls with higher fluorescence signal due to better transfection efficiency (smaller expression plasmids), we usually sorted a 5% population of cells with mean fluorescent intensity (MFI or GeoMean) matched with the top5% MFI of BE3-P2A-EGFP or ABEmax-P2A-EGFP transfected cells from the same day. Please see the provided data for gating strategies in different experimental contexts.

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