nature research

Corresponding author(s):	George M. Church, Jan Gorodkin, Lin Lin, Yonglun Luo
Last updated by author(s):	Mar 7, 2022

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 our Editorial Policies 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	no software was used
Data analysis	FastaQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and fastp (https://github.com/OpenGene/fastp) with default options were used for data quality control and filtering with the default parameters. The pair-end data was assembled using flash (http://www.cbcb.umd.edu/software/flash). BWA-MEM with default options was used to map the assembled data to the designed oligos sequence to preliminarily distinguish the data of each synthetic surrogate site. The pysam module of Python-3.8 was used to split the aligned data according to the site number of the chip. Read filtering was performed with with julia-1.5.3 language. Fisher's exact test and other statistical analysis were performed in R-4.0.3. Visualization was completed by R and excel.

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 and 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 NGS data generated by this study have been shared via the CNGB public data depository with the following accession numbers: CNP0001979, CNP0002648. A complete list of 704 NGS samples were summarized in Supplementary Data 7.

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 sample size calculation was performed.
Data exclusions	To exclusion reads with low synthetic quality during the oligonucleotide synthesis and sites which might affect cell growth, this study used the following filtering steps to exclude low quality sites. (1) Total clean reads less than 32 for both MOCK and SpCas9. (2) percentage indel frequency in MOCK cells above 4%. (3) Fold change of enrichment or depletion between SpCas9 edited and MOCK cells larger than 2 folds.
Replication	The SURRO-seq method has been validated with three libraries.
Randomization	not relevant for the study
Blinding	not relevant for the study

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 and archaeology
[✗]	Animals and other organisms
[✗]	Human research participants
[✗]	Clinical data
[✗]	Dual use research of concern

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)	SpCas9-expressing HEK293T cells were generated by the Luo lab previously. The wild type HEK293T, U2OS, and SKOV-3 were from ATCC. The PC9 cells were purchased from Merck. Human primary fibroblasts were established by the Luo lab.
Authentication	None of the cell lines used were authentificated
Mycoplasma contamination	All cells were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	n.a

Flow Cytometry

Plots

Confirm that:

- [x] The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- [x] 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).
- [x] All plots are contour plots with outliers or pseudocolor plots.
- [x] A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cultured cells were trypsinized and resuspended in PBS. Cell suspensions were analyzed with a NovoCyte Quanteon flow cytometer with 4 lasers.
Instrument	NovoCyte Quanteon flow cytometer with 4 lasers provided by the FACS CORE facility at the Department of Biomedicine, Aarhus University.
Software	Data collection and analysis were performed using the software included in the NovoCyte Quanteon flow cytometer.
Cell population abundance	Flow cytometry was used to quantify the fraction of GFP positive cells. In our experiment, the GFP positive cells are the major population in cells transfected with an EGFP mRNA.
Gating strategy	The major cell population was gated based FSC-A/SSC-A to select the major cell populated. Singlets were gated based on FSC-A and FSC-H. Un-transfected cells were used for setting the gate for GFP negative cells. At least 10,000 cells were collected for each sample.

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