nature portfolio

Corresponding author(s): Rajat Rohatgi and Ganesh V. Pusapati

Last updated by author(s): May 5th, 2026

Reporting Summary

Nature Portfolio 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 Portfolio 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

Sequencing: Illumina NextSeq 550 and MiSeq platforms
Flow cytometry and cell sorting: Sony SH800 Cell Sorter
Immunofluorescence imaging: Leica TCS SP8 confocal microscope
Quantitative real-time PCR: Agilent QuantStudio 5 Real-Time System
Biolayer interferometry: Sartorius Octet RED96 system
Live-Cell Analysis System: Sartorius Incucyte S3

Data analysis

Base editing screen analyses: Custom R scripts and computational pipelines used for processing and statistical analysis of the base-editing screen data were written in R v4.3.2 and are publicly available at https://github.com/RohatgiLab/Murugesh_DC_base_editing_analysis and Zenodo (https://doi.org/10.5281/zenodo.20031065)(murugeshngit-dev 2026). Software used to predict base-editing outcomes is available at https://github.com/RohatgiLab/base-editor-design-tool and Zenodo (https://doi.org/10.5281/zenodo.19957892)(murugeshngit-dev 2026).

Visualization of quantitative data: GraphPad Prism (version 10)
Structural figures: UCSF ChimeraX (version 1.8rc202405140121). PDB codes for structural figures are provided in the respective figure legends.
Immunoblotting processing: Adobe PhotoShop (for chemiluminescence data) and Fiji (for LI-COR data).
Immunofluorescence quantifications: CellProfiler (v4
Flow cytometry: FlowJo (version 10)
Cell proliferation: Incucyte AI confluence analysis tool

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All raw and processed sequencing data supporting the CRISPR-based screening and subsequent analyses are available in the Sequence Read Archive (SRA) repository under BioProject accession number PRJNA1344315. The specific read counts and processed data files derived from the initial base editing screen are listed in Supplementary Table 1. All source data and corresponding raw quantification files used for the validation of sgRNAs in eHAP1-7TS and HEK293T cell lines are provided in Supplementary Table 3, accompanying the paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

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

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 statistical methods were used for sample size determination. Unless otherwise indicated, all experiments were performed with a minimum of three (n≥3) independent biological replicates. Technical replicates were not counted as independent samples for statistical analysis.
Data exclusions	No data were excluded from analyses.
Replication	All attempts to repeat experiments were successful. All experiments were repeated on separate days to verify reproducibility with the exception of the initial validation screen. Due to the large scale of hits (370 sgRNAs), flow cytometry measurements were conducted as a single experiment in both eHAP1-7TS and HEK293T-7TS reporter cell lines. The selection of sgRNAs for subsequent, more detailed characterization was based on the criterion that a given sgRNA showed the same phenotype (either a reduction or an elevation in fluorescence) in both cell lines. sgRNAs that met this criterion were then subjected to detailed characterization in at least three independent experiments.
Randomization	No randomization was performed.
Blinding	No blinding was performed.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

The following primary antibodies were used for immunoblotting and immunofluorescence: β-catenin (BD Biosciences, Cat#610154, Clone: 14/Beta-Catenin, 1:3,000 for immunoblotting), Non-phospho (Active) β-Catenin (Ser33/37/Thr41) (Cell Signaling Technology, Cat#8814, Clone: D13A1, 1:1,000 for immunoblotting, 1:200 for immunofluorescence), Phospho β-Catenin (Ser33/37) (Cell Signaling Technology, Cat#2009, 1:1000 for immunoblotting), AXIN1 (Cell Signaling Technology, Cat#2074, Clone: C95H11, 1:1,000 for immunoblotting), APC (Millipore-Sigma, Cat#OP44, Clone: FE9, 1:200 for immunoblotting), LEF1 (Cell Signaling Technology, Cat#2230, Clone: C12A5, 1:1,000 for immunoblotting), TCF3 (Cell Signaling Technology, Cat#2883, Clone: D15G11, 1:1,000 for immunoblotting), TCF4 (Cell Signaling Technology, Cat#2569, Clone: C48H11, 1:1,000 for immunoblotting), E-cadherin (Cell Signaling Technology, Cat#3195, Clone: 24E10, 1:1,000 for immunoblotting), GAPDH (Proteintech, Cat#60004-1-Ig, Clone: 1E6D9, 1:10,000 for immunoblotting), α-Tubulin (Proteintech, Cat#66031-1-Ig, Clone: 1E4C11, 1:10,000 for immunoblotting), and GST (Bethyl Laboratories, Cat#A190-122A, 1:1000 for immunoblotting). The following secondary antibodies were used: Peroxidase AffiniPure Donkey Anti-Mouse IgG (H+L) (Jackson ImmunoResearch; Cat#715-035-151, 1:10,000 for immunoblotting), Peroxidase AffiniPure Donkey Anti-Rabbit IgG (H+L) (Jackson ImmunoResearch; Cat#711-035-152, 1:10,000 for immunoblotting), IRDye 800CW Donkey Anti-Mouse IgG (LICORbio; Cat#926-32212, 1:5,000 for immunoblotting), IRDye 800CW Donkey Anti-Rabbit IgG (LICORbio; Cat#926-32213, 1:5,000 for immunoblotting), and Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 488 (Thermo Fisher Scientific; Cat#A-21202, 1:1,000 for immunofluorescence).

Validation

All commercially available antibodies were validated previously by the manufacturer.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)

HAP1, a near haploid human cell line and eHAP1, a fully haploid human cell line, were gifts from Jan Carette, Stanford University. HEK293T, RKO, DLD-1, SW480, LoVo, and HT29 cell lines were obtained from American Type Culture Collection (ATCC).

Authentication

HAP1 and eHAP1 were These cell lines were authenticated by morphology and ploidy analysis by flow cytometry with propidium iodide staining. The identity of HEK293T, RKO, DLD-1, SW480, LoVo, and HT29 cell lines was authenticated by the vendor using STR profiling.

Mycoplasma contamination

All cell lines were confirmed to be mycoplasma-free using an isothermal PCR-based assay (InvivoGen, Cat#rep-mys-10).

Commonly misidentified lines
(See ICLAC register)

None used.

Plants

Seed stocks

n/a

Novel plant genotypes

n/a

Authentication

n/a

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	eHAP1-7TS, HEK293T-7TS, RKO APC 1320*-7TG, and DLD-1-7TG cells were harvested by trypsinization. Cells were resuspended in growth medium and transfered to a FACS tube containing a nylon mesh cell strainer. Tubes were placed on ice for no more than 2 h.
Instrument	Sony SH800 Cell Sorter
Software	FlowJo (version 10)
Cell population abundance	To ensure data quality, cell population abundance was assessed for consistency. Any sample with significantly reduced event counts was re-processed and analyzed in a separate experiment on a different day. Samples that reproducibly showed low abundance in the repeat experiment were classified as outliers and excluded from the dataset.
Gating strategy	Debris was excluded and the main cell population was identified based on its forward (FSC-A) and side (SSC-A) scatter properties. Singlets were then gated using FSC-A vs. FSC-W. Fluorescence (mScarlet or GFP) was quantified from this gated, single-cell population. Positive gates were established based on untreated cells.

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