nature portfolio

Corresponding author(s): B. P. Kleinstiver

Last updated by author(s): 2025/03/06

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 data was acquired via next-generation sequencing (Illumina MiSeq and Illumina NextSeq)
Data analysis	Data was plotted using GraphPad Prism 10 (v10.0.1). Next-generation sequencing data was analyzed using CRISPResso2. Figures were generated using Adobe Illustrator (v28.6). Enzymes were homology modeled in Coot (v0.9.8.93) and visualized using ChimeraX (v1.8). Custom software used in this study include: Guide-seq-2 analysis software (https://github.com/tsailabSJ/guideseq/tree/V2) HT-PAMDA analysis software (https://github.com/kleinstiverlab/HT-PAMDA) Github links for code developed in this study: PAMmla ML models and ISDE (https://github.com/RachelSilverstein/PAMmla)) AND (https://pammla.streamlit.app/) Analysis of multiplex sequencing of SpCas9 variants (https://github.com/RachelSilverstein/multiplex_seq_analysis)

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

Next-generation sequencing results are available through the NCBI sequence read archive (SRA) under PRJNA1169103. PAMmla predictions for all 64 million SpCas9(6AA) enzymes can be viewed through an online webtool (https://pammla.streamlit.app/). UniRef100 dataset used to generate multiple sequence alignments for natural sequence models can be downloaded at https://www.uniprot.org/uniref/.

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	*Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design; whether sex and/or gender was determined based on self-reporting or assigned and methods used.* *Provide in the source data disaggregated sex and gender data, where this information has been collected, and if consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected.* *Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.*
Reporting on race, ethnicity, or other socially relevant groupings	*Please specify the socially constructed or socially relevant categorization variable(s) used in your manuscript and explain why they were used. Please note that such variables should not be used as proxies for other socially constructed/relevant variables (for example, race or ethnicity should not be used as a proxy for socioeconomic status).* *Provide clear definitions of the relevant terms used, how they were provided (by the participants/respondents, the researchers, or third parties), and the method(s) used to classify people into the different categories (e.g. self-report, census or administrative data, social media data, etc.)* *Please provide details about how you controlled for confounding variables in your analyses.*
Population characteristics	*Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write "See above."*
Recruitment	*Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.*
Ethics oversight	*Identify the organization(s) that approved the study protocol.*

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	Sample sizes for mouse studies were chosen based on prior retinal gene editing studies (PMC6980783), balancing statistical power with ethical considerations to minimize animal use in alignment with the 3Rs principle.
Data exclusions	Mice were excluded from analysis if subretinal injection was unsuccessful, as evidenced by ocular damage or the absence of detectable BFP expression in the retina at the time of dissection or if less than 1,000 BFP+ sorted cells could be recovered during FACS sorting.
Replication	Experimental data are expressed as the mean and standard deviation of at least two or three independent replicates, with replicate experiments performed on separate days when possible. Attempts at replication were successful, in that we believe that the results from our experiments reasonably approximate the mean.
Randomization	No randomization was performed in our study as appropriate control samples were used to establish baseline/reference activities. Randomization is not relevant to our study.

Blinding	No blinding was performed in our study as no subjective measurements/ measurements dependent on researcher observation were involved in 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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HEK293T cells (ATCC); Epstein Barr virus thransformed B cell lines (BLCs) were established at the NIH (Following consent uder NIH protocol 05-I-0213) from a patient with Chronic Granulomatous Disease harboring the CYBB T362I mutation
Authentication	cell line identities were confirmed by STR profiling (ATCC)
Mycoplasma contamination	Media supernatant was analyzed biweekly for the presence of Mycoplasma and all tests were negative for contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	Humanized WT-hRHO-GFP and P23H-hRHO-RFP mice were gifted from Dr. Theodore G. Wensel. The two strains were crossbred to generate the heterozygous hRHO-WT-GFP/hRHO-P23H-RFP mouse line. All experiments were performed on P0-P2 heterozygous pups. Equal numbers of male and female mice were used. Mice were housed under a 12-hour light/dark cycle at an ambient temperature of 20-22°C and relative humidity of 40-60%.
Wild animals	The study did not involve wild animals
Reporting on sex	Equal numbers of male and female mice were used.
Field-collected samples	The study did not involve field-collected samples.
Ethics oversight	Our animal study followed the tenets of the Association for Research in Vision and Opthalmology Statement for the Use of Animals in Opthalmic and Vision Research and the guidelines of the Massachusetts Eye and Ear for Animal Care and Use (under IACUC protocol number 2021N000059)

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

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.