nature portfolio

Corresponding author(s): James M. Carothers, Jesse G. Zalatan

Last updated by author(s): Jun 28, 2024

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	For RNA folding, energetic parameters were generated using the RNAfold, RNAeval, RNAduplex, and Findpath programs from the ViennaRNA Package version 2.3.5. Sequences of full scRNAs were input to a custom Python 2.7 script that used those programs to return the energetic parameters. Custom Python code to analyze input RNA and generate the energetic parameters described in this work is available on GitHub (https://github.com/carothersresearch/gRNA_screen_docker OR https://doi.org/10.5281/zenodo.12558212). Jupyter notebooks to view and reproduce the ART results from this paper are available on GitHub (https://github.com/carothersresearch/art_lnt OR https://doi.org/10.5281/zenodo.12559439) Existing tools for CRISPR guide design and evaluation were implemented using the CRISPOR webserver (http://crispor.tefor.net/), with the default settings for "No Genome" and Protospacer Adjacent Motif (PAM) set to "20bp-NGG - SpCas9, SpCas9-HF1, eSpCas9 1.1". Each 20 bp target was evaluated using the "predicted guide efficiency" output. Plate reader data was collected using BioTek Gen5 2.07.17. Flow cytometry data was collected using MACSQuantify 2.8.
Data analysis	Data were plotted and statistically analyzed using Graphpad Prism 8.4.3.686 and Microsoft Excel 15.17. Median fluorescence in flow cytometry data was calculated using FlowJo 10.0.7.

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

Data supporting the findings of this work are available within the paper and its Supplementary Information files. Source data are provided with this paper as a Source Data file. Supplementary information available for this paper includes the Supplementary Information file, Supplementary Table 2, Supplementary Table 3, Supplementary Table 7, GenBank Sequences file, and Source Data file.

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 were determined following standard practice in the field (Qi et al, 2013 doi: 10.1016/j.cell.2013.02.022; Bikard et al., 2013 doi: 10.1093/nar/gkt520; Dong et al., 2018 doi: 10.1038/s41467-018-04901-6). No statistical tests were used to predetermine sample size.
Data exclusions	In Figure 4b, one replicate was excluded from strain #9 (high-J3, low-J5, high-J6) because the culture failed to grow.
Replication	Experiments were performed in triplicate to ensure reproducibility between biological replicates. The exception is the full-library LNT analysis in Figures 6b, 6c, and S11, because of the low-throughput nature of the HPLC analysis. A smaller experiment analyzed by HPLC (Fig 6e), however, showed high agreement between biological replicates. Biological replication was performed by picking a single separate colony from plasmid transformation plates for each replicate. For HPLC samples, technical variation in chromatogram y-axis scaling was normalized by quantifying an endogenous peak unrelated to LNT production. Replication was found to be successful in all replicated experiments.
Randomization	To inoculate biological replicate cultures, single colonies were chosen randomly from plasmid transformation plates. Experimental groups, e.g.

Randomization	different expression conditions, were determined by genotype, and therefore allocation of samples into these groups was irrelevant, since each transformation plate contained only colonies of the same genotype.
Blinding	The investigators were not blinded as the data collected in this work did not require a subjective analysis by the experimenters.

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

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.*

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	E. coli cells from overnight cultures were diluted 1:50 in DPBS before analysis on the flow cytometer.
Instrument	MACSQuant VYB flow cytometer
Software	Collection: MACSQuantify 2.8; Analysis: FlowJo 10.0.7
Cell population abundance	Bacterial cell populations were gated for single cells as described below. 10,000 events were collected per sample.
Gating strategy	A figure exemplifying the gating strategy is available in our prior work (doi: 10.1038/s41467-018-04901-6). A side scatter threshold trigger (SSC-H) was applied to enrich for single cells. A narrow gate along the diagonal line on the SSC-H vs SSC-A plot was selected to exclude the events where multiple cells were grouped together. Within the selected population, events that appeared on the edges of the FSC-A vs. SSC-A plot and the fluorescence histogram were excluded.

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