nature portfolio

Corresponding author(s): Jason W. Chin

Last updated by author(s): 17/05/2023

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	Illumnina sequencer (MiSeq / HiSeq 2500/ NextSeq 2000) were used to acquire all sequencing data in this study.
Data analysis	Graphpad Prism version 9 was used to generate all bar graphs. FCS Express Flow Cytometry software (version 7) was used for FACS data analysis. For sequence alignments custom scripts were used. Code used to generate recoding landscape is available at https://github.com/JWChin-Lab/recoding-landscape-generator; for sequence data alignment and analysis at https://github.com/JWChin-Lab/NGS-analysis; for analysis of BASIS constructs at https://github.com/JWChin-Lab/BASIS-sequence-analysis; for automated liquid handling for next generation sequencing library prep at https://github.com/JWChin-Lab/NGS-sample-prep.

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 availability

The sequences and design details used in this study are available in the Supplementary Data. Supplementary Data 1 provides all spacer sequences used in CONEXER and BASIS experiments; Supplementary Data 2 lists all nucleotide sequences of oligonucleotides, plasmids and BACs used in this study; Supplementary Data 3

provides the GenBank file of spacer plasmid pKW3 to express universal spacer set 1; Supplementary Data 4 provides the GenBank file of spacer plasmid pKW3 to express universal spacer set 2; Supplementary Data 5 provides the GenBank file of a general CONEXER BAC design with spacer sequences of universal spacer set 1; Supplementary Data 6 provides the GenBank file of a general CONEXER BAC design with spacer sequences of universal spacer set 2; Supplementary Data 7 provides the GenBank file of plasmid pLF118 for λ-red recombineering; Supplementary Data 8 provides the GenBank file of CFTR BAC01; Supplementary Data 9 provides the GenBank file of CFTR BAC02; Supplementary Data 10 provides the GenBank file of CFTR BAC03; Supplementary Data 11 provides detailed results for sequence verification, listing all variants and their classification as called for the final CFTR assembly; Supplementary Data 12 lists all raw sequencing data, including their NCBI SRA accession numbers, as deposited under BioProject PRJNA962525; Supplementary Data 13 provides the GenBank file of plasmid pFR015 used for retron-editing; Supplementary Data 14 provides the GenBank file of helper plasmid pFR156 for retron-editing; Supplementary Data 15 provides the GenBank file of plasmid pHBA008 which served as template to amplify BASIS components for human BAC adaptation; Supplementary Data 16 provides the GenBank file of plasmid pHBA010 which served as template to amplify BASIS components for human BAC adaptation; Supplementary Data 17 provides the GenBank file of initial assembly acceptor plasmid pHBA031 for the 1.1 Mbp BASIS assembly; Supplementary Data 18 provides detailed results for sequence verification, listing all true positive variants as called for the final 1.1 Mbp BASIS assembly; Supplementary Data 19 provides detailed results for sequence verification, listing all variants as called for the final 1.1 Mbp BASIS assembly and categorised as either likely false positive or false positive; Supplementary Data 20 provides the GenBank file of plasmid pSP43 with spacer sequences for gene knock-out by CRISPR/Cas9-mediated cleavage and λ-red recombineering.

All other datasets generated and/or analysed in this study are available from the corresponding author upon reasonable request. All materials (Supplementary Data 3, 4, 7, 8, 9, 10, 13, 14, 15, 16, 17, 20) from this study are available from the corresponding author upon reasonable request

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. These are biochemical experiments, not animal experiments, with no individuals to sample. Sample size is not a relevant parameter but number of replicates. Number of replicates is indicated. Further, the variation in the assays used is small and we are interested in large effects, the sample sizes used, as indicated in the manuscript, were deemed appropriate.
Data exclusions	There are no data exclusions.
Replication	The exact number of replicates is stated in the relevant figure legend/methods. All attempts at replication were successful.
Randomization	No randomization. This is not relevant because the samples form defined groups. Further, these are biochemical experiments, where different components need to be added to different reactions. A single investigator needs to perform defined distinct and skilled operations on different samples to make the experiment meaningful and therefore randomization is not applicable.
Blinding	These are biochemical experiments, where different components need to be added to different reactions. A single investigator needs to perform defined distinct and skilled operations on different samples to make the experiment meaningful and therefore blinding is not meaningful and applicable.

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)	HEK293 cells were used (commercially available from European Collection of Authenticated Cell Cultures (ECACC)).
Authentication	HEK293 cell line was authenticated by ECACC using STR DNA profiling analysis to rule out intra- or inter-species contamination.
Mycoplasma contamination	All cell lines used were tested negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

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 trypsinized and resuspended in PBS.
Instrument	FACS Fortessa - BD LSRFortessa™ Cell Analyzer
Software	FCS express
Cell population abundance	Total cells sorted: CFTR transfection: 1'906'511 / neg. control: 806'185. The fraction of cells that are GFP positive is indicated in the Figure.
Gating strategy	Untransfected (GFP negative) cells were used for gating.

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