nature portfolio

Corresponding author(s): Anthony McDowell Mustoe

Last updated by author(s): Jun 28, 2025

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 collected using either an in-house MiSeq instrument or a NovaSeq instrument from the Genomic and RNA Profiling Core at Baylor College of Medicine. Electropherogram was collected using Agilent 4200 Tapestation. Agarose gel image was collected using Bio-Rad ChemiDoc MP Imaging System. Western blot image was collected using Bio-Rad ChemiDoc XRS System.
Data analysis	Custom codes are available in the Dao_MPRAsplicing folder at https://github.com/MustoeLab/Publications, archive of Github repository is available at https://doi.org/10.5281/zenodo.15700943 Tools used: Python (v3.8.5), samtools (v1.3.1), bbmap (v39.19), bbmerge (v39.19), biopython (v1.72), edlib (v.1.3.8.post2), pysam (v.0.16.0.1), pandas (v.2.0.3), numpy (v.1.24.4), matplotlib (v.3.7.3), seaborn (v.0.12.2), statsmodel (v.0.14.1), spliceAI (v.1.3.1) (https://github.com/Illumina/SpliceAI), Pangolin (v. 1.0.2) (https://github.com/tkzeng/Pangolin), and XSTREME (https://meme-suite.org/meme/tools/xstreme). Quantification of band intensity in agarose gel was done with ImageJ (v1.54p). Quantification of band intensity in western blot was done with Image Lab Software (Bio-Rad ChemiDoc XRS System).

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

Sequencing data is available in the NCBI SRA under BioProject accession code PRJNA1116243 (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1116243). Processed sequencing data (expression and splicing quantification) are available at https://github.com/MustoeLab/Publications the under MIT license and have been archived at https://doi.org/10.5281/zenodo.15700943. Full plasmid maps for the 3 representative plasmids are available as Supplementary Data 1-3. Source Data are provided with this 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	Not applicable
Reporting on race, ethnicity, or other socially relevant groupings	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

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 sample size was calculated. For MPRA experiments, we performed two or three biological replicates, which is standard for published MPRA studies.
Data exclusions	All MPRA data were used, and no exclusion was done prior to analysis.
Replication	To assess reproducibility, we calculated Pearson correlation for reporter expression and for splicing fraction between biological replicates. In all experiments, we obtained high Pearson correlations, indicating high reproducibility.
Randomization	No randomization was performed. We controlled for covariates by performing multiple biological replicates and ensuring our measurements were reproducible.
Blinding	No blinding was performed. Blinding was unnecessary for our MPRA experiments because expression and splicing fraction was quantified using software and thus not affected by knowledge of sample identities.

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

Living Colors® A.v. Monoclonal Antibody (JL-8) at 1:5000 dilution; Clonetech; catalog number 632381
β-Actin (8H10D10) Mouse mAb (HRP Conjugate) at 1:10000 dilution; Cell Signaling; catalog number 12261

Validation

Living Colors® A.v. Monoclonal Antibody (JL-8) was validated by immunoblotting by the manufacturer (https://www.takarabio.com/documents/Certificate%20of%20Analysis/632380/632380-632381-070313.pdf?srsltid=AfmBOoo2Kyzw2wvO-rc54VWbdF_Gwhju11Uma5kuntz33rUFkll6IsGb)
β-Actin (8H10D10) Mouse mAb (HRP Conjugate) was validated by immunoblotting by the manufacturer (https://www.cellsignal.com/products/antibody-conjugates/b-actin-8h10d10-mouse-mab-hrp-conjugate/12262)

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	HeLa cells were obtained from Baylor College of Medicine Tissue Culture Core. HEK293 (R78007), SH-S5Y5 (CRL-2266), U87 MG (HTB-14) were obtained from the American Type Culture Collection (ATCC).
Authentication	Cells were authenticated using M.D.Anderson Cytogenetics and Cell Authentication Core.
Mycoplasma contamination	All cell lines were tested negative for mycoplasma contamination using LookOut Mycoplasma PCR Detection Kit.
Commonly misidentified lines (See ICLAC register)	No cell lines used are commonly misidentified lines

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable