

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 (<i>n</i>) 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 <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
☒	☐ 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. <i>F</i> , <i>t</i> , <i>r</i>) with confidence intervals, effect sizes, degrees of freedom and <i>P</i> value noted <i>Give P values as exact values whenever suitable.</i>
☒	☐ 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 <i>d</i> , Pearson's <i>r</i>), 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	Collection of the data included next generation sequencing on an Illumina NovaSeq X plus (performed by Admera Health, NJ, USA) using the corresponding software (v 1.2.1).
Data analysis	Data Analysis included mapping RNA sequencing reads using STAR (v2.7.11b) in multiplexed format using custom made python and bash scripts. All the code used for the mapping and data processing is available on GitHub (https://github.com/shakedshanas/Splicing-MPRA-data-processing). Subsequent steps were performed using custom python scripts (python version 3.12.2 using numpy 1.26.4, pandas 2.2.2, scipy 1.13.1, statsmodels 0.14.2, scikit-posthocs 0.11.4., matplotlib 3.9.2, seaborn 0.13.2). This code is available upon request.

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](#)

Raw and processed data have been deposited in ArrayExpress (accession number E-MTAB-16553).

The mapping and processing pipeline used here is available on github: <https://github.com/shakedshanas/Splicing-MPRA-data-processing> and described there and in the methods section.

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](https://www.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

Altogether 9,608 reporter sequences were tested in 11 different promoter constructs and 8 different conditions. Sample sizes represent subsets of these libraries, depending on the specific analysis, and are indicated in the Figure legends. As most analysis use the entire available data (after filtering), the sample size typically was >5000 in order to account for the complexity of regulatory interactions and to detect potentially smaller effect sizes.

Data exclusions

Data exclusion and filtering partially depended on the specific analysis and is described in the Methods and in the Results section at the appropriate positions. Generally, sequencing reads were filtered for correct mapping of at least 110 out of 150 nucleotides in both the forward and reverse read, as well as based on the number of mismatches and the extent of soft clipping during STAR mapping. Generally, isoforms supported by a single read in any sample were discarded. Additional filtering steps vary and are described at the relevant place in the manuscript.

Replication

Reproducibility of the results was confirmed on the one hand by having two independent biological replicates (starting from library cloning), as well as technical replicates in the form of identical sequence variants with different barcode sequences who share all processing and analysis steps.

Randomization

For each experiments, all samples were processed together in an anonymized manner in order to mitigate batch effects.

Blinding

The investigators carrying out the individual steps were not aware of the experimental groups.

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)	HEK293 cells were a gift from Eran Segal (and originally obtained from ATCC).
Authentication	Authentication was done based on morphological features.
Mycoplasma contamination	Cells were tested for mycoplasma and found negative.
Commonly misidentified lines (See ICLAC register)	n/a

Plants

Seed stocks	<i>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.</i>
Novel plant genotypes	<i>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.</i>
Authentication	<i>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, mosaicism, off-target gene editing) were examined.</i>

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	HEK293 cells were washed with PBS, trypsinized and resuspended in DMEM+10%FBS by pipetting.
Instrument	A NovoCyte (Agilent) equipped with two lasers (488 nm and 561 nm) and an automatic sampler was used.
Software	Gating and acquisition was done using the NovoCyte software. FCS files were exported and processed using custom python scripts.
Cell population abundance	No sorting was performed. The abundance of the analyzed population was influenced by the transfection efficiency and the expression rate (representative examples are shown in Supplementary Figure 17).
Gating strategy	Gating based on forward and side scatter was performed to ensure selection of single, intact cells for analysis (representative examples are shown in Supplementary Figure 17). An mCherry positive population was then selected (FI in the 561 channel >10 ⁴) to ensure equal representation of all promoter variants, including lowly expressing ones.
☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.	