nature portfolio

Corresponding author(s): Y. Jessie Zhang

Last updated by author(s): 8-20-2024

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	Commercial softwares equipped by LSM710 confocal microscope (Zeiss), NIS-Elements Viewer (Nikon), Plate reader (Tecan), AB Voyager-DE PRO MALDI-TOF (Brunker Corporation), NovaSeq 6000 (Illumina)
Data analysis	Orgin Pro 9.1, ZEN 2010, ImageJ, GraphPad Prism 8.0, RStudio 4.0.5, trim-galore v.0.6.7, bowtie2 v.2.4.5, rMATS-turbo v.4.1.2, SpliceTools, deepTools2 v.3.5.1, samtools 1.9, bwtool 1.0, MACS2 v.2.2.7.1, hisat2 v.2.2.1, custom algorithms, https://github.com/tailana703/CTD_length_PolII_pausing_alternative_splicing

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

Source data are provided with this paper. The sequencing data generated in this study have been deposited in GEO under the accession number: GSE252261

(https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252261), GSE252258 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252258) (ChIP-Seq, HEK293T), GSE252260 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252260) (RNA-Seq, HEK293T).

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	n/a
Reporting on race, ethnicity, or other socially relevant groupings	n/a
Population characteristics	n/a
Recruitment	n/a
Ethics oversight	n/a

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	According to the experience to choose an adequate pool for turbidity assay, fluorescence polarization, microscopy, chip-seq and RNA-seq experiments. No statistical method was used to predetermine sample size.
Data exclusions	No exclusion
Replication	For each representative image/data, at least 1 other additional experiments were successfully repeated.
Randomization	Randomization was not performed.
Blinding	The investigators were not blinded during experiments and outcome assessment.

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	Stated in the Methods section for the source, catalog number and clone number, and dilution rate of antibodies in distinct applications.

Validation	Validation of antibodies is provided by individual commercial providers including Santa Cruz, Abcam, Millipore, Proteintech, Active Motif in their user manuals and websites. Stated in the Methods section for validation of antibodies.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Cell lines (U2OS, HEK 293T) were from ATCC.
Authentication	The cell lines were not authenticated.
Mycoplasma contamination	All cell lines were tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No cell lines used in this study were found in the database of commonly misidentified cell lines that is maintained by ICLAC and NCBI Biosample.

Plants

Seed stocks	n/a
Novel plant genotypes	n/a
Authentication	n/a

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?&acc=GSE252261; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252258; https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252260
Files in database submission	HEK 293T, RPB1_52xCTD, replicate 1; HEK 293T, RPB1_52xCTD, replicate 2; HEK 293T, RPB1_26xCTD, replicate 1; HEK 293T, RPB1_26xCTD, replicate 2; HEK 293T, RPB1_T4E_all, replicate 1; HEK 293T, RPB1_T4E_all, replicate 2; HEK 293T, RPB1_T4E_sp, replicate 1; HEK 293T, RPB1_T4E_sp, replicate 2; HEK 293T, RPB1_S7E_sp, replicate 1; HEK 293T, RPB1_S7E_sp, replicate 2; HEK 293T, IgG, replicate 1; HEK 293T, IgG, replicate 2
Genome browser session (e.g. UCSC)	hg38

Methodology

Replicates	All of the experiments contain 2 replicates for each sample.
Sequencing depth	Each sample has 150 bp length of reads (paired-end library).
Antibodies	GFP antibody (proteintech, 50430-2-AP)
Peak calling parameters	Map with Bowtie2 (default parameters with paired-end library, index was built using reference genome hg38 in .fasta format); Used MACS2 on filtered aligned reads (mapq > 10): --format BAMPE --gsize '2700000000' --broad --broad-cutoff='0.005' --keep-dup '1' --d-min 20 --buffer-size 100000 --pvalue '0.005' --nomodel --extsize '200' --shift '0' 2>&1 > macs2_stderr Then, peaks were imported into R for DiffBind.
Data quality	Initial quality assessment showed high library complexity with low level of duplication. Adapter sequences and low-quality read ends were trimmed off by TrimGalore! v.0.6.7 with default parameters.
Software	trim-galore v.0.6.7, bowtie2 v.2.4.5, deepTools2 v.3.5.1, samtools 1.9, bwtool 1.0, MACS2 v.2.2.7.1, RStudio 4.0.5