nature research

Corresponding author(s): Eric S. Lander and Mathias Munschauer

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Spectrum Mill v6.01 pre-release, NIS elements, CytExpert.
Data analysis	Graphpad Prism 7, FlowJo, Microsoft Excel, R studio, MATLAB (version R2017b), FISH-QUANT, Cell Profiler (version 2.2.0 and version 3.0.0), BEDTools, Integrative Genomics Viewer (version 2.3.26), Igvtools 2.3, STAR v2.5.2a, TopHat version 2.0.8, MACS2, Picard's MarkDuplicates, Bowtie 2, DESeq2, SUPPA2, diffSplice, kallisto, Stringtie v1.3.3b, MAFFT.

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/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Sequencing data associated with Fig. 2, 3, and Extended Data Fig. 3, 4, 5 is available at the Gene Expression Omnibus under the accession number GSE114953.

The original mass spectra associated with Fig. 1, 3, and Extended Data Fig. 4 may be downloaded from MassIVE (http://massive.ucsd.edu) using the identifier: MSV000012345. The data is directly accessible via ftp://massive.ucsd.edu/MSV000012345.

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical measures were used to predetermine sample size. We typically performed at least 3 biological replicates for each experiment, unless otherwise noted in the figure legends. Quantitative mass spectrometry experiments using isobaric mass tag labeling include 2 replicate experiments in each 4plex or 6plex mass tag cassette as is common practice.
Data exclusions	Mass spectrometry data were filtered for common laboratory contaminants. We further required a minimal number of unique peptides as defined in the Methods. Otherwise no data were excluded.
Replication	As reported in the figure legends, main text and Methods, the findings were reliably reproduced.
Randomization	This study uses a candidate-based approach to dissect molecular interactions and biochemical mechanisms. No randomization was required because the results of biochemical measurements or sequencing of nucleic acid libraries are not affected by sample randomization.
Blinding	This study uses a candidate-based approach to dissect molecular interactions and biochemical mechanisms. No blinding was required because the results of biochemical measurements or sequencing of nucleic acid libraries are not affected by knowledge of sample identities. DNA combing was performed by scientists at specialized service facilities and sample identities were not disclosed to the person performing the experiment or data analysis.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

RBMX – Cell Signaling (D7C2V) #14794, Lot #1, 2, (IP (6 ug/ml), western blot (1:1000), proximity ligation assay (1:250), immunofluorescence (1:250))
RBMX – Santa Cruz Biotechnology (G17) #sc-14581, Lot #K0614, (IP (6 ug/ml) and western blot (1:1000))
ALYREF – Bethyl #A302-892A, Lot #1, (IP (6 ug/ml))
ALYREF – Santa Cruz Biotechnology (11G5) #sc-32311, Lot # J1016, (western blot (1:1000))
PUM1 – Bethyl #A302-577A, Lot #1, (IP (6 ug/ml))
V5 – Abcam (SV5-Pk1) #ab27671, Lot #GR242588-4, (IP (6ug/ml))
V5 – Abcam #ab9116, Lot #GR256657-12, (western blot (1:1000))
Flag – Cell Signaling (9A3) #8146S, Lot #3, (proximity ligation assay (1:250))
TOP1 – Bethyl #A302-589A, Lot #1, (IP (6 ug/ml), proximity ligation assay (1:250))
TOP1 – Bethyl #A302-590A, Lot #3, (western blot 1:1000)
TOP1 – Santa Cruz Biotechnology (C-21) #sc-32736, Lot #K0515, (western blot (1:1000))
TOP1 – Thermo Fisher Scientific (24.5) #435900, Lot #QC215904, (proximity ligation assay (1:250))
CDC5L – Bethyl #A301-681A, Lot #1, (IP (6 ug/ml))
CDC5L – Santa Cruz Biotechnology (2136C1a) #sc-81220, Lot #C0917, (western blot (1:1000))
PRPF19 – Bethyl #A300-101A, Lot #1, (western blot (1:500))

ACTB – Abcam #ab8226 (western blot (1:1000))
BRCA1 – Bethyl #A301-378A, Lot #2, (western blot (1:1000))
BRCA2 – Bethyl #A300-005A (western blot (1:1000))
Alpha-Tubulin–FITC – Sigma Aldrich #F2168 (immunofluorescence (1:1000))
Anti-Centromere antibodies (ACA) – Antibodies Incorporated #15-234-0001 (immunofluorescence (1:200))
Normal Rabbit IgG - Cell signaling #2729, Lot #8, (IP (6 ug/ml))

Validation

All antibodies were validated by western blot. Commercial antibodies were also validated by the manufactures as indicated on their web sites.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HCT116 cell line was acquired from ATCC.
Authentication	Cell lines were not authenticated.
Mycoplasma contamination	Cell lines tested negative for mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

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	A detailed description of the sample preparation procedure is given in the Methods sections "Cell cycle analysis".
Instrument	Beckman Coulter Cytoflex S
Software	CytExpert, FlowJo.
Cell population abundance	No post-sort fractions were collected.
Gating strategy	Cell cycle analysis: We applied forward and side scatter parameters (FSC, SSC) to exclude cell debris and doublets as exemplified in Supplementary Figure 1. Typically, 3-4% of events were excluded to remove events that could not be assigned to cells in G1-, S-, and G2-M-phase. The fraction of cells in each phase of the cell cycle was then quantified using indicated gates (contour plot).

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