nature research

Corresponding author(s): Joanna Wysocka and Michael Bassik

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

For the screen, two independent replicate screens were performed, which are sufficient for screening technologies. See Methods section, 'Genome-wide screen in K562 cells', 'Secondary screen in K562 cells' and 'Genome-wide screen and secondary screen in HeLa cells'.

2. Data exclusions

Describe any data exclusions.

No data was excluded.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Once experiments and procedures were fully optimized, all attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

K562 KO clones were allocated into experimental groups based on their genotype (Extended Data Figure 4c). For smFISH experiments, individual cells were allocated into experimental groups based on their genotype (Figure 2b and Extended Data Figure 4d).

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Blinding was performed during the smFISH data analyses (Extended Data Figure 4e), where B.G. who analyzed smFISH images was blinded to all conditions.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	For Images analyses we used MATLAB 2016b (MathWorks). Flowjo 9.9 was used for flow cytometry analyses. For statistical analysis we used R 3.3.2. For ChIP-seq genomic alignments we used bowtie2 v.2.2.9, peak calls with MACS2 v2.1.1.20160309, IGV_2.3.92 and IGB 9.0.0 for visualization, bedtools v2.17.0 and GNU awk 4.1.3 for overlap statistics and genome interval manipulation. For humans hg38 reference genome was used, for mouse mm10. RNAseq alignments were performed with hisat2 v2..0.5, followed by stringtie v 1.3.3b and featureCounts v1.4.6-p2, further analysis was performed with Bioconductor 3.4 and DESeq2 1.14.1, human genocde 25 transcript models were used.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	1) Rabbit MORC2 antibody (A300-149A, Bethyl Laboratories), validated by vendor, and used in previous literature. 2) Rabbit MPP8 antibody (16796-1-AP, Protein Technologies Inc), validated by vendor, and used in previous literature. 3) Rabbit TASOR antibody (HPA006735, Atlas Antibodies), validated by vendor, and used in previous literature. 3) Mouse anti-LINE-1 ORF1p antibody (MABC1152, Millipore), validated by vendor, and used in previous literature. 4) Rabbit HSP90 (C45G5, Cell Signalling, #4877), Extensively used in the literature. 5) Beta actin antibody (ab49900, Abcam), Extensively used in the literature. 6) Histone H3 (tri-methyl K9) antibody (ab8898, Abcam), validated by vendor, and used in previous literature. 7) RNA Pol II (Santa Cruz Biotechnology, N-20 sc-899), validated by vendor, and used in previous literature.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Cell lines are from commercial sources. HeLa and K562: ATCC Human Embryonic Stem Cells, H9: WiCell Mouse Embryonic Stem Cells, ES-E14TG2a: ATCC
b. Describe the method of cell line authentication used.	Cell lines were authenticated by the vendor. All cells were obtained from comerical sources. HeLa, K562 and mESC (ATCC). Human Embryonic Stem Cells H9 (WiCell).
c. Report whether the cell lines were tested for mycoplasma contamination.	Cell cultures were routinely tested and found negative for mycoplasma infection (MycoAlert, Lonza).
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None of the cell lines used in this study are in the database of commonly misidentified cell lines.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	No animals were used in this study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	This study did not involve human research participants.

nature research

Corresponding author(s): Joanna Wysocka, Michael Bassik

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

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

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

2. Provide all necessary reviewer access links.
The entry may remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=ojerwuukpzsbjsb&acc=GSE95374

3. Provide a list of all files available in the database submission.

GSM2509455
ChIP:MORC2_Cell:WT_rep1
GSM2509456
ChIP:MORC2_Cell:WT_rep2
GSM2509457
ChIP:MORC2_Cell:MORC2-KO_rep1
GSM2509458
ChIP:MORC2_Cell:MORC2-KO_rep2
GSM2509459
ChIP:MORC2_Cell:MPP8-KO_rep1
GSM2509460
ChIP:MORC2_Cell:MPP8-KO_rep2
GSM2509461
ChIP:MORC2_Cell:TASOR-KO_rep1
GSM2509462
ChIP:MORC2_Cell:TASOR-KO_rep2
GSM2509463
ChIP:MPP8_Cell:WT_rep1
GSM2509464
ChIP:MPP8_Cell:WT_rep2
GSM2509465
ChIP:MPP8_Cell:MORC2-KO_rep1
GSM2509466
ChIP:MPP8_Cell:MORC2-KO_rep2
GSM2509467
ChIP:MPP8_Cell:MPP8-KO_rep1
GSM2509468
ChIP:MPP8_Cell:MPP8-KO_rep2
GSM2509469
ChIP:MPP8_Cell:TASOR-KO_rep1
GSM2509470
ChIP:MPP8_Cell:TASOR-KO_rep2
GSM2509471
ChIP:TASOR_Cell:WT_rep1
GSM2509472
ChIP:TASOR_Cell:WT_rep2
GSM2509473
ChIP:TASOR_Cell:MORC2-KO_rep1
GSM2509474
ChIP:TASOR_Cell:MORC2-KO_rep2

June 2017

GSM2509475
ChIP:TASOR_Cell:MPP8-KO_rep1
GSM2509476
ChIP:TASOR_Cell:MPP8-KO_rep2
GSM2509477
ChIP:TASOR_Cell:TASOR-KO_rep1
GSM2509478
ChIP:TASOR_Cell:TASOR-KO_rep2
GSM2509479
ChIP:Input(MORC2, MPP8, TASOR)_Cell:WT_rep1
GSM2509480
ChIP:Input(MORC2, MPP8, TASOR)_Cell:WT_rep2
GSM2509481
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MORC2-KO_rep1
GSM2509482
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MORC2-KO_rep2
GSM2509483
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MPP8-KO_rep1
GSM2509484
ChIP:Input(MORC2, MPP8, TASOR)_Cell:MPP8-KO_rep2
GSM2509485
ChIP:Input(MORC2, MPP8, TASOR)_Cell:TASOR-KO_rep1
GSM2509486
ChIP:Input(MORC2, MPP8, TASOR)_Cell:TASOR-KO_rep2
GSM2509487
ChIP:H3K9me3_Cell:WT_rep1
GSM2509488
ChIP:H3K9me3_Cell:WT_rep2
GSM2509489
ChIP:H3K9me3_Cell:MORC2-KO_rep1
GSM2509490
ChIP:H3K9me3_Cell:MORC2-KO_rep2
GSM2509491
ChIP:H3K9me3_Cell:TASOR-KO_rep1
GSM2509492
ChIP:H3K9me3_Cell:TASOR-KO_rep2
GSM2509493
ChIP:Input(H3K9me3)_Cell:WT_rep1
GSM2509494
ChIP:Input(H3K9me3)_Cell:WT_rep2
GSM2509495
ChIP:Input(H3K9me3)_Cell:MORC2-KO_rep1
GSM2509496
ChIP:Input(H3K9me3)_Cell:MORC2-KO_rep2
GSM2509497
ChIP:Input(H3K9me3)_Cell:TASOR-KO_rep1
GSM2509498
ChIP:Input(H3K9me3)_Cell:TASOR-KO_rep2
GSM2509503
ChIP:H3K9me3_Cell:WT_rep3
GSM2509504
ChIP:H3K9me3_Cell:WT_rep4
GSM2509505
ChIP:H3K9me3_Cell:MPP8-KO_rep1
GSM2509506
ChIP:H3K9me3_Cell:MPP8-KO_rep2
GSM2509507
ChIP:Input(H3K9me3)_Cell:WT_rep1
GSM2509508
ChIP:Input(H3K9me3)_Cell:WT_rep2

GSM2509509
ChIP:Input(H3K9me3)_Cell:MPP8-KO_rep1
GSM2509510
ChIP:Input(H3K9me3)_Cell:MPP8-KO_rep2
GSM2789802
ChIP:PolII_Cell_K562:WT_rep1
GSM2789803
ChIP:PolII_Cell_K562:WT_rep2
GSM2789804
ChIP:PolII_Cell_K562:MORC2-KO_rep1
GSM2789805
ChIP:PolII_Cell_K562:MORC2-KO_rep2
GSM2789806
ChIP:PolII_Cell_K562:MPP8-KO_rep1
GSM2789807
ChIP:PolII_Cell_K562:MPP8-KO_rep2
GSM2789808
ChIP:PolII_Cell_K562:TASOR-KO_rep1
GSM2789809
ChIP:PolII_Cell_K562:TASOR-KO_rep2
GSM2789810
ChIP:Input_Cell_K562:WT_rep1
GSM2789811
ChIP:Input_Cell_K562:WT_rep2
GSM2789812
ChIP:MPP8_Cell_hESC:WT_rep1
GSM2789813
ChIP:MPP8_Cell_hESC:WT_rep2
GSM2789814
ChIP:MORC2_Cell_hESC:WT_rep1
GSM2789815
ChIP:MORC2_Cell_hESC:WT_rep2
GSM2789816
ChIP:Input_Cell_hESC:WT_rep1
GSM2789817
ChIP:Input_Cell_hESC:WT_rep2
GSM2789818
ChIP:MPP8_Cell_mESC:WT_rep1
GSM2789819
ChIP:MPP8_Cell_mESC:WT_rep2
GSM2789820
ChIP:Input_Cell_mESC:WT_rep1
GSM2789821
ChIP:Input_Cell_mESC:WT_rep2
MORC2KO1_CHIP_INPUT_1.bw
MORC2KO1_CHIP_INPUT.bw
MORC2KO1_CHIP_K9me3.bw
MORC2KO1_CHIP_MORC2_1.bw
MORC2KO1_CHIP_MPP8_1.bw
MORC2KO1_CHIP_TASOR_1.bw
MORC2KO2_CHIP_INPUT_1.bw
MORC2KO2_CHIP_INPUT.bw
MORC2KO2_CHIP_K9me3.bw
MORC2KO2_CHIP_MORC2_1.bw
MORC2KO2_CHIP_MPP8_1.bw
MORC2KO2_CHIP_TASOR_1.bw
MPP8_1.bw
MPP8_2.bw
MPP8KO1_CHIP_INPUT_1.bw
MPP8KO1_CHIP_INPUT_K9_1.bw

MPP8KO1_CHIP_K9me3.bw
MPP8KO1_CHIP_MORC2_1.bw
MPP8KO1_CHIP_MPP8_1.bw
MPP8KO1_CHIP_TASOR_1.bw
MPP8KO2_CHIP_INPUT_1.bw
MPP8KO2_CHIP_INPUT_K9.bw
MPP8KO2_CHIP_K9me3.bw
MPP8KO2_CHIP_MORC2_1.bw
MPP8KO2_CHIP_MPP8_1.bw
MPP8KO2_CHIP_TASOR_1.bw
SAFE_1.bw
SAFE_2.bw
TASOR_1.bw
TASOR_2.bw
TASORKO1_CHIP_INPUT_1.bw
TASORKO1_CHIP_INPUT.bw
TASORKO1_CHIP_K9me3.bw
TASORKO1_CHIP_MORC2_1.bw
TASORKO1_CHIP_MPP8_1.bw
TASORKO1_CHIP_TASOR_1.bw
TASORKO2_CHIP_INPUT_1.bw
TASORKO2_CHIP_INPUT.bw
TASORKO2_CHIP_K9me3.bw
TASORKO2_CHIP_MORC2_1.bw
TASORKO2_CHIP_MPP8_1.bw
TASORKO2_CHIP_TASOR_1.bw
WT1_CHIP_INPUT_1.bw
WT1_CHIP_INPUT.bw
WT1_CHIP_INPUT_K9.bw
WT1_CHIP_K9me3.bw
WT1_CHIP_MORC2_1.bw
WT1_CHIP_MPP8_1.bw
WT1_CHIP_TASOR_1.bw
WT2_CHIP_INPUT_1.bw
WT2_CHIP_INPUT.bw
WT2_CHIP_INPUT_K9.bw
WT2_CHIP_K9me3.bw
WT2_CHIP_MORC2_1.bw
WT2_CHIP_MPP8_1.bw
WT2_CHIP_TASOR_1.bw
hESC_Input1.bw
hESC_Input2.bw
hESC_MORC2_ChIP1.BW
hESC_MORC2_ChIP2.BW
hESC_MPP8_ChIP1.BW
hESC_MPP8_ChIP2.BW
INPUT_1.bw
INPUT_2.bw
INPUT_mESC1.bw
INPUT_mESC2.bw
MPP8_mESC1.bw
MPP8_mESC2.bw
PolII_MORC2_KO1.bw
PolII_MORC2_KO2.bw
PolII_MPP8_KO1.bw
PolII_MPP8_KO2.bw
PolII_TASOR_KO1.bw
PolII_TASOR_KO2.bw
PolII_WT1.bw

	PolII_WT2.bw
4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).	n/a

Methodological details

5. Describe the experimental replicates.	ChIP experiments (MORC2, MPP8, TASOR, H3K9me3 and RNA PolII) were performed in two biological replicates each, with indicated antibodies. Peaks were extensively validate using ChIP-qPCR.
6. Describe the sequencing depth for each experiment.	To amplify each library we used quantitative PCR (qPCR) to endure that all libraries were amplified similarly and avoid bottlenecking of the libraries. ChIP-Seqs are pair ended, 75 bp was the read length. On average, each ChIP-seq sample contain ~40 million reads, with above 70-80% alignment.
7. Describe the antibodies used for the ChIP-seq experiments.	Rabbit MORC2 antibody (A300-149A, Bethyl Laboratories), Rabbit MPP8 antibody (16796-1-AP, Protein Technologies Inc), Rabbit TASOR antibody (HPA006735, Atlas Antibodies), Histone H3 (tri-methyl K9) antibody (ab8898, Abcam) and RNA Pol II (Santa Cruz Biotechnology, N-20 sc-899) were used in ChIP experiments.
8. Describe the peak calling parameters.	Pair-end reads were trimmed with cutadapt (-m 50 -q 10) and aligned with bowtie2 (version 2.2.9, --no-mixed --no-discordant --end-to-end -maxins 500) to the hg38 reference genome. ChIP peak calls were performed with macs2 callpeak using default settings, except for --broad flag. Background files were either ChIP input sequencing or ChIPseq from knockout cell lines for factor ChIPped. For final list of sites MACS2 peak calls were merged, combined with 2x amount of shuffled decoy sites and read coverage for each sequencing file was obtained using bedtool coverage. Combined coverage matrix was subjected to DESeq2 procedure to reject false positives from MACS2
9. Describe the methods used to ensure data quality.	Visualization tracks were generated with bedtools genomecov (-bg -scale) with scaling factor being 10^6/number aligned reads and converted to bigWig with bedGraphToBigWig (Kent tools). BigWigs were plotted with IGV browser. Individual alignments were inspected with IGB browser. Heatmaps were generated by intersecting bam alignment files with intervals of interest (bedtools v2.25.0), followed by tabulation of the distances of the reads relative to the center of the interval and scaling to account for total aligned read numbers (10^6/number aligned). Heatmaps were plotted using a custom R function. Aggregate plots were generated by averaging rows of the heatmap matrix. ChIP-seq repetitive sequence relationship analysis. Repeat masker was intersected with ChIP-seq peak calls to classify each masker entry as MPP8 bound, MORC2-bound or unbound. Enriched families of repeats were identified with R fisher.test() followed by FDR correction with qvalue(). Distribution of sizes of occupied vs non-occupied L1 was plotted using R density() with sizes being taken from repeat masker. ks.test() was used to reject null hypothesis that distribution of sizes for bound and unbound L1s is the same. To investigate relationship between L1 age, length and occupancy, logistic regression was performed with R glm() engine. Quantitative analysis of H3K9me3 changes was performed by first identifying regions of significant enrichment in each sample relative to corresponding input sample (macs2 callpeak), merging the intervals into a common superset. This superset was joined with a decoy randomized set of intervals, twice the size of actual experimental interval set, with the same size distribution (bedtools shuffle). Next the read coverage was

determined for each sample (bedtools coverage) and regions with significant change together with fold changes were identified using DESeq2 analysis paradigm. H3K9me3 regions were classified into bound vs unbound by performing intersect with MORC2 and MPP8 ChIP peak calls.

10. Describe the software used to collect and analyze the ChIP-seq data.

For ChIP-seq, we used standard and available softwares: Bowtie, MACS2, Bedtools. Details are provided in the Methods sections.
All sequencing samples reported have been deposited at GEO under the accession number: GSE95374. Detailed Data and further code information are available on request from the authors.

nature research

Corresponding author(s): Joanna Wysocka and Michael Bassik

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. 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).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Live cells were sorted. No staining involved.
6. Identify the instrument used for data collection.	BD LSR Fortessa
7. Describe the software used to collect and analyze the flow cytometry data.	BD Diva for collection and FlowJo for analysis
8. Describe the abundance of the relevant cell populations within post-sort fractions.	The abundance of transposition positive cells is generally low. ~300,000 gated events were collected for each sample to determine GFP(+) fractions, with target of at least 200 positive cells collected.
9. Describe the gating strategy used.	Cells were gated for live/dead and doublet exclusion using FSC and SSC channels, then cells were gated for presence of mCherry signal (reporting on presence of gRNA). Events passing above gating strategy were classified as positive or negative based on SSC and GFP channel signals.

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