nature research

Corresponding author(s): Rosenfeld, G Michael
Li, Wenbo

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	StepOnePlus™ Software v2.3 PXi 6 GeneSys software
Data analysis	Home-made codes are deposited in GitHub (https://github.com/wblilab-uth/ERR-project and https://github.com/zdz-lab/ERR). GraphPad Prism version 9 Juicebox 1.11.08 samtools 1.9 Bowtie 2 version 2.2.6 Bedtools version 2.25.0 HOMER version 4.8.2 MACS2 version 2.1.2 Illumina Basecalling Pipeline CASAVA 1.8.4 R/Bioconductor package version 3.6.1 The usages of software were described in the method section. All codes are available upon request. Data and Code availability information has been provided in the manuscript as a separate section.

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

Code and Data availability statement was included in Methods Section.
Data and Code Availability: The datasets generated during and/or analyzed during the current study are available in the GEO (GSE115604) and SRA (PRJNA412021). For analyses of GTEx data and cancer mutations, codes are deposited in GitHub (https://github.com/wblilab-uth/ERR-project and https://github.com/zdz-lab/ERR). Other analyses of ChIP-Seq peaks, 4C-Seq or gene expression in the current study used standard bioinformatics tools and codes, which can be available upon request.
The association with traits and diseases were identified using genome wide association studies in the NHGRI-EBI GWAS Catalog(https://www.ebi.ac.uk/gwas/) and GWASdb v2(https://doi.org/10.1093/nar/gkv1317).
The GTEx data used for the analyses described in this manuscript were obtained from dbGap accession: phs000424.v7.p2.c1 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000424.v7.p2).

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	Sample size calculation was not performed. We determined the number of samples in each experiment as commonly accepted standards in the field. For qPCR and ChIP-qPCR, two or three biological replicates, each of three technical replicates was conducted. For key NGS data such as GRO-seq and 4C-seq, two to three biological replicates were conducted. We described these in the Figure Legends, Methods Section and Extended Data table 5.
Data exclusions	No data were excluded.
Replication	The data in this paper is highly replicable, as two laboratories conducted many experiments separately and reproduced all conclusions. We generated many promoter knockout isogenic cell clones to exclude single-cell clone-specific phenomenon. We performed two biological replicates for Extended Data Fig. 1e; 2a, b, d, e, g, h; 3a-c; 4d, e; 6f; 8e; 9d; 11d three biological replicates for Fig. 2c, d, f, g; 3b, d, f; 4e; Extended Data Fig. 1a, d; 4c, f; 9e, f; four biological replicates for Fig. 3g; 4f GRO-seq of RAD21 KD: 3 replicates, GRO-seq of TFF1pKO cells; 2 replicates, SMC1a, H3K4me3, H3K4me1 ChIP: n=1 but confirmed in -/+ E2 signals. 4C-seq for TFF1e, TFF3p, KCNK5e, P2RY2e, NUCKS1p, RAB7L1p: 2 replicates
Randomization	This study does not involve work that requires random allocation. Most of the experiments are comparisons of isogenic control vs mutant cell line or control cells vs cells under specific treatment.
Blinding	The investigators were not blinded to data analysis (e.g. gene expression qPCR, GRO-seq or ChIP-seq). But for this current study, blinding is not relevant.

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

antibody information was provided in Method Section
ERα Santa Cruz Sc543 ChIP and ChIP-seq
p300 Santa Cruz Sc585 ChIP and ChIP-seq
RNA Pol II Santa Cruz Sc-899 ChIP and ChIP-seq
H3K4me3 Abcam Ab8580 ChIP-seq
H3K4me1 Abcam Ab8895 ChIP-seq
SMC1a Bethyl A300-055A ChIP-seq
BrU agarose beads Santa Cruz Sc-32323 ac GRO-seq
V5 Invitrogen R960-25 WB
Cas9 Proteintech 26758-I-AP WB
CTCF Millipore 07-729 ChIP and WB

Validation

All the antibodies used in this paper have been widely used and cited. The ChIP-seq pattern (e.g. H3K4me3 that mostly locates to promoters) also indicated that the the antibody is specific.
ERα Santa Cruz https://www.scbt.com/p/eralpha-antibody-hc-20
p300 Santa Cruz https://www.scbt.com/p/p300-antibody-c-20
RNA Pol II Santa Cruz http://datasheets.scbt.com/sc-899.pdf
H3K4me3 Abcam https://www.abcam.com/histone-h3-tri-methyl-k4-antibody-chip-grade-ab8580.html
H3K4me1 Abcam https://www.abcam.com/histone-h3-mono-methyl-k4-antibody-chip-grade-ab8895.html
SMC1a Bethyl https://www.bethyl.com/product/A300-055A/SMC1+Antibody
BrU agarose beads Santa Cruz https://www.scbt.com/p/brdu-antibody-iib5
V5 Invitrogen https://www.thermofisher.com/antibody/product/V5-Tag-Antibody-Monoclonal/R960-25
Cas9 Proteintech https://www.ptglab.com/products/Cas9-Antibody-26758-1-AP.htm
CTCF Millipore https://www.emdmillipore.com/US/en/product/Anti-CTCF-Antibody,MM_NF-07-729

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Cell lines, MCF-7, HEK293T, were from ATCC originally.
Two hiPSC (iPSCORE_2_11 and iPSCORE_19_1) were obtained from iPSCORE at UCSD and were differentiated to NPCs.
These are also described in the Method section.

Authentication

We purchase new batch of cells every 1-2 years from ATCC to ensure the authentication.

Mycoplasma contamination

Described in Methods. MCF-7 and HEK293T were examined for Mycoplasma contamination every 6-12 months.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in our study.

Methodology

Replicates	For key experiments, two to three biological replicates were conducted (e.g. GRO-seq for siRAD21 or promoter KO cells.)
Sequencing depth	ChIP-seq-SMC1a_EtOH 28,677,353 ChIP-seq-SMC1a_E2 28,940,838 ChIP-seq-H3K4me3_siCTL + E2 11,576,209 ChIP-seq-H3K4me3_siRAD21 + E2 15,215,510 ChIP-seq-H3K4me1_siCTL + E2 16,680,182 ChIP-seq-H3K4me1_siRAD21 + E2 23,303,290 ChIP-seq-H3K4me3_WT MCF7 + E2 17,872,206 ChIP-seq-H3K4me3_TFF1p-KO MCF7 + E2 23,200,670 GRO-seq-nascent RNA_siCTL + EtOH (repeat1) 6,863,150 GRO-seq-nascent RNA_siCTL + E2 (repeat1) 7,667,228 GRO-seq-nascent RNA_siRAD21 + EtOH (repeat1) 9,471,922 GRO-seq-nascent RNA_siRAD21 + E2 (repeat1) 6,926,569 GRO-seq-nascent RNA_siCTL + EtOH (repeat2) 12,624,935 GRO-seq-nascent RNA_siCTL + E2 (repeat2) 13,276,916 GRO-seq-nascent RNA_siRAD21 + EtOH (repeat2) 15,855,814 GRO-seq-nascent RNA_siRAD21 + E2 (repeat2) 16,884,136 GRO-seq-nascent RNA_siCTL + EtOH (repeat3) 8,771,184 GRO-seq-nascent RNA_siCTL + E2 (repeat3) 7,815,340 GRO-seq-nascent RNA_siRAD21 + EtOH(repeat3) 5,762,301 GRO-seq-nascent RNA_siRAD21 + E2 (repeat3) 15,631,681 GRO-seq-nascent RNA_Wildtype cell EtOH 9,675,923 GRO-seq-nascent RNA_Wildtype cell E2 11,844,838 GRO-seq-nascent RNA_TFF1p-KO (clone28) EtOH 11,633,638 GRO-seq-nascent RNA_TFF1p-KO (clone28) E2 13,282,685 GRO-seq-nascent RNA_TFF1p-KO (clone25) EtOH 20,015,942 GRO-seq-nascent RNA_TFF1p-KO (clone25) E2 17,259,248 4C-seq_TFF1e_viewpoint_wt_R1 4,155,458 4C-seq_TFF1e_viewpoint_TFF1pKO_R1 3,460,929 4C-seq_TFF1e_viewpoint_wt_R2 1,237,120 4C-seq_TFF1e_viewpoint_TFF1pKO_R2 648,531 4C-seq_TFF1e_viewpoint_wt_R3 6,651,744 4C-seq_TFF1e_viewpoint_TFF1pKO_R3 7,251,685 4C-seq_TFF3p_viewpoint_wt_R1 683,138 4C-seq_TFF3p_viewpoint_TFF1pKO_R1 189,726 4C-seq_TFF3p_viewpoint_wt_R2 2,996,729 4C-seq_TFF3p_viewpoint_TFF1pKO_R2 2,800,472 4C-seq_TFF3p_viewpoint_TFF1pKO_R3 5,424,601 4C-seq_TFF1e_viewpoint_mcf7_sictl_E21h_R1 4,111,503 4C-seq_TFF1e_viewpoint_mcf7_sirad21_E21h_R1 3,990,561 4C-seq_TFF1e_viewpoint_mcf7_sictl_E21h_R2 32,084,133 4C-seq_TFF1e_viewpoint_mcf7_sirad21_E21h_R2 26,919,259 4C-seq_P2RY2e_viewpoint_mcf7_sictl_E21h_R1 6,068,548 4C-seq_P2RY2e_viewpoint_mcf7_sirad21_E21h_R1 5,769,008 4C-seq_P2RY2e_viewpoint_mcf7_sictl_E21h_R2 583,828 4C-seq_P2RY2e_viewpoint_mcf7_sirad21_E21h_R2 754,506 4C-seq_KCNK5e_viewpoint_mcf7_sictl_E21h_R1 1,007,371 4C-seq_KCNK5e_viewpoint_mcf7_sirad21_E21h_R1 2,090,372 4C-seq_KCNK5e_viewpoint_mcf7_sictl_E21h_R2 185,863 4C-seq_KCNK5e_viewpoint_mcf7_sirad21_E21h_R2 82,854 4C-seq_TERTp_viewpoint_293t_WT 2,756,635 4C-seq_TERTp_viewpoint_293t_CLPTM1Lp-KO 2,156,282 4C-seq_PAX5p_viewpoint_293t_WT 5,421,873 4C-seq_PAX5p_viewpoint_293t_ZCCHC7p-KO 3,231,924 4C-seq_nucks1p_viewpoint_alt_R1 34,713 4C-seq_nucks1p_viewpoint_ref_R1 33,468 4C-seq_nucks1p_viewpoint_alt_R2 37,309 4C-seq_nucks1p_viewpoint_ref_R2 49,169 4C-seq_RAB7L1p_viewpoint_alt_R1 18,388 4C-seq_RAB7L1p_viewpoint_ref_R1 14,017 4C-seq_RAB7L1p_viewpoint_alt_R2 253,592 4C-seq_RAB7L1p_viewpoint_ref_R2 262,205
Antibodies	Described in Methods Section H3K4me3 Abcam Ab8580 H3K4me1 Abcam Ab8895 SMC1a Bethyl A300-055A BrU agarose beads Santa Cruz Sc-32323 ac
Peak calling parameters	MACS2 was used for peak calling using standard setting with q-value 0.01.

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.*	GEO/GSE115604: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115604 and SRA (PRJNA412021): https://trace.ncbi.nlm.nih.gov/Traces/sra/?study=SRP118797
Files in database submission	Provided in Extended Data Table 5
Genome browser session (e.g. UCSC)	https://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=wenbo%20li&hgS_otherUserSessionName=Oh_Paper_UCSC_2018
Data quality	ChIP-Seq reads were subjected to FASTQC for quality check.
Software	Samtools 1.9 Bowtie 2 version 2.2.6 Bedtools version 2.25.0 HOMER version4.8.2 R/Bioconductor package version 3.6.1.