nature research

Corresponding author(s): Ana Losada, Ana Cuadrado, Marc A. Martí-Renom

☐ 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.

   n/a

2. Data exclusions

   Describe any data exclusions.

   n/a

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   For ChIP-sequencing we have confirmed the distribution of cohesin SA1 and cohesin SA2 by ChIP with an antibody against SMC1. In addition, we have confirmed our results in three different human cell lines (MCF10A, HCAEC and HMEC). For ChIP-seq from control, SA1 and SA2 depleted MCF10A cells, two replicates were performed and sequenced for each antibody.

   Several positions (common and SA2-only) were validated by means of ChIP-qPCR (3 technical replicates each).

   To assess Wapl/SA2 ratio in different positions by ChIPqPCR we performed at least three experimental replicates (each with three technical replicates).

   For re-ChIP experiment we ensured the reliability of the colocalization by reciprocal ChIP of cohesin SA1 and SA2 subunits. Additionally, we included IgG and SMC1 controls.

   For Hi-C we performed two replicates (two independent libraries) per condition. Even if in the main figure the analyses were performed combining reads from both replicates, replicates were also analyzed independently and gave similar results (Supplementary Fig.5 and 6).

   For proteomic analysis, a single immunoprecipitation experiment per antibody was performed with two technical replicates. Non-immune IgG was used as control.

4. Randomization

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

   n/a

5. Blinding

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

   n/a

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.	All the software used to analyze the data is specified in the online Methods section and in the ChIP-seq report and is publicly available.

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.	Custom-made antibodies are available in reasonable amounts upon request

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	Antibodies used for ChIP are described in the ChIP-seq report. A rat monoclonal antibody against SA1 was generated using a 233-aa long N-terminal fragment as antigen and validated for immunoblotting in extracts from WT and SA1 KO MEFs.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	Primary human cell lines were purchased from Lonza . MCF10A cell line was obtained from Dr. Quintela (CNIO, Madrid)
b. Describe the method of cell line authentication used.	For authentication of MCF10A cell line a karyotype analysis was performed by the Cytogenetics Unit at CNIO
c. Report whether the cell lines were tested for mycoplasma contamination.	Commercial cell lines were provided mycoplasm-free. MCF10A cells were periodically tested for mycoplasm (last test was performed right before expansion) with the GEN-PROBE MTC-NI rapid detection system.
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.	Not applicable

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.

For laboratory animals, report species, strain, sex and age OR for animals observed in or captured from the field, report species, sex and age where possible OR state that no animals were used.

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.

Provide all relevant information on human research participants, such as age, gender, genotypic information, past and current diagnosis and treatment categories, etc. OR state that the study did not involve human research participants.

nature research

Corresponding author(s): Ana Losada, Ana Cuadrado, Marc A. Martí-Renom

☐ 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.

Link to the GEO submission page: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101921
access token for reviewers: ybatoyqivfmzzsf

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

SA1_HMEC_ChIP-seq
SA2_HMEC_ChIP-seq
SMC1_HMEC_ChIP-seq
Input_HMEC_ChIP-seq
SA1_MCF10A_ChIP-seq
SA2_MCF10A_ChIP-seq
SMC1_MCF10A_ChIP-seq
ZMYM2_MCF10A_ChIP-seq
Input_MCF10A_ChIP-seq
SA1_HCAEC_ChIP-seq
SA2_HCAEC_ChIP-seq
SMC1_HCAEC_ChIP-seq
Input_HCAEC_ChIP-seq
SA1_MCF10A_ChIP-seq_Control_Rep_1
SA1_MCF10A_ChIP-seq_Control_Rep_2
SA2_MCF10A_ChIP-seq_Control_Rep_1
SA2_MCF10A_ChIP-seq_Control_Rep_2
SA1_MCF10A_ChIP-seq_siSA1_Rep_1
SA1_MCF10A_ChIP-seq_siSA1_Rep_2
SA2_MCF10A_ChIP-seq_siSA1_Rep_1
SA2_MCF10A_ChIP-seq_siSA1_Rep_2
SA1_MCF10A_ChIP-seq_siSA2_Rep_1
SA1_MCF10A_ChIP-seq_siSA2_Rep_2
SA2_MCF10A_ChIP-seq_siSA2_Rep_1
SA2_MCF10A_ChIP-seq_siSA2_Rep_2
INPUT_MCF10A_ChIP-seq_Control
INPUT_MCF10A_ChIP-seq_siSA1
INPUT_MCF10A_ChIP-seq_siSA2

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

http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=Dinamica%20cromosomica&hgS_otherUserSessionName=Reviewers%20Session

Methodological details

5. Describe the experimental replicates.

We have performed single replicas of each ChIP-seq with >40 million reads depth sequencing for most conditions except for SMC1 ChIP in MCF10A (2 replicates) and for SA1 and SA2 ChIP in control, siSA1 and siSA2 treated

MCF10A cells (2 replicates per condition). ChIP have been performed in three different cell lines (HMEC, MCF10A and HCAEC) with three different cohesin antibodies (SMC1, SA1 and SA2).

6. Describe the sequencing depth for each experiment.

Library preparation and PCR conditions: Fragmented DNA samples, quantified by fluorometry were processed through subsequent enzymatic treatments of end-repair, dA-tailing, and ligation to adapters with "NEBNext Ultra II DNA Library Prep Kit for Illumina" from New England BioLabs (catalog # E7645). Adapter-ligated libraries were completed by limited-cycle PCR and extracted with a [single] double-sided SPRI size selection. Median fragment size is 300 bp from which 120 bp correspond to adaptor sequences. Libraries were applied to an Illumina flow cell for cluster generation and sequenced on an Illumina instrument (see below) by following manufacturer's protocols.

HMEC and HCAEC (SA1, SA2, SMC1 and Input): 8 ng per sample. 13 cycles.
MCF10A (SA1, SA2, SMC1 and Input): ~5.5ng per sample. 13 cycles.
MCF10A, Zmym2: 6,4 ng. 15 cycles
For ChIP-seq performed in siC, siSA1 and siSA2 MCF10A cells (marked with asterisk below) 10 ng were used per sample, except for the following samples where 5 ng were used:
SA2_MCF10A_ChIP-seq_Control_Rep_1
SA2_MCF10A_ChIP-seq_siSA1_Rep_1
SA2_MCF10A_ChIP-seq_siSA1_Rep_2
11 cycles of PCR were performed in all samples, except for the sample 'SA2_MCF10A_ChIP-seq_siSA2_Rep_2' –in which 13 cycles were performed.
For Re-ChIP, 1ng of DNA was used and 18 cycles of PCR were performed

Experiment	No of reads \ uniquely mapped
SA1_HMEC_ChIP-seq	97,741,618 78,687,032
SA2_HMEC_ChIP-seq	98,226,569 82,156,467
SMC1_HMEC_ChIP-seq	95,778,603 79,606,174
Input_HMEC_ChIP-seq	32,277,764 22,041,565
SA1_MCF10A_ChIP-seq	61,973,461 39,702,285
SA2_MCF10A_ChIP-seq	73,440,077 42,638,750
SMC1_MCF10A_ChIP-seq	92,003,268 50,267,535
ZMYM2_MCF10A_ChIP-seq	45,156,594 20,857,810
Input_MCF10A_ChIP-seq	24,028,387 22,436,324
SA1_HCAEC_ChIP-seq	73,350,000 64,838,899
SA2_HCAEC_ChIP-seq	67,000,000 61,110,324
SMC1_HCAEC_ChIP-seq	67,813,000 59,521,632
Input_HCAEC_ChIP-seq	66,200,000 61,267,006
SA1_MCF10A_ChIP-seq_Control_Rep_1	34,918,353 22,086,186 *
SA1_MCF10A_ChIP-seq_Control_Rep_2	39,279,433 24,942,790 *
SA2_MCF10A_ChIP-seq_Control_Rep_1	34,995,732 21,790,656 *
SA2_MCF10A_ChIP-seq_Control_Rep_2	35,646,335 22,717,313 *
SA1_MCF10A_ChIP-seq_siSA1_Rep_1	35,498,232 23,027,702 *
SA1_MCF10A_ChIP-seq_siSA1_Rep_2	33,085,555 20,729,535 *
SA2_MCF10A_ChIP-seq_siSA1_Rep_1	37,670,724 23,870,685 *
SA2_MCF10A_ChIP-seq_siSA1_Rep_2	32,109,651 20,374,192 *
SA1_MCF10A_ChIP-seq_siSA2_Rep_1	33,615,601 21,306,063 *
SA1_MCF10A_ChIP-seq_siSA2_Rep_2	35,651,442 22,825,770 *
SA2_MCF10A_ChIP-seq_siSA2_Rep_1	33,757,359 20,848,771 *
SA2_MCF10A_ChIP-seq_siSA2_Rep_2	65,758,370 28,872,448 *
INPUT_MCF10A_ChIP-seq_siSA2	46,072,845 30,803,519 *
INPUT_MCF10A_ChIP-seq_siSA1	47,124,889 32,254,566 *
INPUT_MCF10A_ChIP-seq_Control	47,185,555 31,020,560 *
Re_ChiP_SA2-IgG-MCF10A	4,607,572 1,028,764
Re_ChiP_SA2-SA1-MCF10A	5,943,225 923,656

*In the case of calibrated ChIP-seq, number of reads in the right column corresponds to reads obtained after separating reads coming from human chromatin from reads coming from mouse chromatin in the mapping step and after processing.
Length of the (single end) reads was 75bp for calibrated ChIPseq and 50 bp for all the rest.

7. Describe the antibodies used for the ChIP-seq experiments.

Home made Cohesin complex antibodies against SMC1, SA1, SA2 and Wapl have been validated using the following strategy:

1. Western blotting in cohesin knocked down cells
To assess the specificity of these antibodies we performed western blot in different cell types depleted from each individual subunit as follows:
• SMC1 specificity was tested in MCF10A cells depleted from SMC1 with siRNA on target SMART pool L-006833 (Dharmacon).
• SA1 specificity was tested in MCF10A cells depleted from SA1 with siRNA on target SMART pool L-010638 (Dharmacon) as well as in SA1KO MEFs
• SA2 specificity was tested in MCF10A cells depleted from SA2 with siRNA on target SMART pool L-010638 (Dharmacon) and MEFs depleted of SA2 with siGENOME SMARTpool M-057033.
• Wapl specificity was tested in mES cells depleted from Wapl with siRNA siGENOME SMART pool M-047528 (Dharmacon)
2. Immunoprecipitation
We successfully performed immunoprecipitation experiments in human and mouse cell lines to assess the ability of the above-mentioned antibodies to recognize and bind its target protein in the context of the cohesin complex.
3. ChIP
• We performed SA1 ChIP-seq in SA1 KO MEFs as described (2). Our data shows a very reduced number of cohesin SA1 peaks (about 600) when compared with those recovered in the wt MEFs (about 26,000).
• To validate the reliability of the positions recovered with cohesin antibodies, we overlap the signals obtained for different subunits. The high degree (typically, higher than 80%) of similarity between SA1 or SA2 and the common subunit SMC1 is an indicator of the efficiency and specificity of the antibodies under discussion.

Zmym2 [a generous gift from H. Yu (UT Southwestern, US)] antibody specificity has been validated by means of immunoprecipitation experiments performed with different members of the Co-REST complex (3). We performed the following additional validations:

1. Western blotting in knockdown cells: To assess the specificity of the Zmym2 antibody, we performed western blotting in mES cells depleted from Zmym2 protein by means of the siRNA on target SMART pool L-064538 (Dharmacon).
2. ChIP-seq: We confirmed by ChiP-qPCR the ability of Zmym2 antibody to recognize some of the already described Zmym2 binding sites in U2OS cells using a FLAG-tagged version of Zmym2 protein (4).

References
1. S. Remeseiro et al., Cohesin-SA1 deficiency drives aneuploidy and tumourigenesis in mice due to impaired replication of telomeres. EMBO J. 31, 2076–2089 (2012).
2. S. Remeseiro, A. Cuadrado, G. Gómez-López, D. G. Pisano, A. Losada, A unique role of cohesin-SA1 in gene regulation and development. EMBO J. 31, 2090–2102 (2012).
3. C. B. Gocke, H. Yu, ZNF198 stabilizes the LSD1-CoREST-HDAC1 complex on chromatin through its MYM-type zinc fingers. PLoS ONE. 3, e3255

(2008).
4. E. Aguilar-Martínez et al., Screen for multi-SUMO–binding proteins reveals a multi-SIM–binding mechanism for recruitment of the transcriptional regulator ZMYM2 to chromatin. PNAS 112(35), E4854–E4863 (2015)

8. Describe the peak calling parameters.

ChIP-seq and control reads were aligned to the hg19 genome assembly using bwa (version 0.6.1-r104) under default parameters. In the case of experiments in calibrated ChIP-seq performed in siC, siSA1 and siSA2 MCF10A cells (marked with asterisk) reads were aligned using bowtie2 (version 2.3.3.1). Peak calling was performed using macs2 (version 2.1.1.20160309) setting following parameters: '-q 0.05' (0.01 in the case of HMEC experiments) and '-extsize (value obtained from macs2 predicted step)' and using input as the control.

9. Describe the methods used to ensure data quality.

Raw read files were assessed using fastqc prior to processing. Unmapped reads were removed using samtools (version 1.3.1) running 'samtools view -F 4'. Reads were sorted and replicates removed using picardtools (version 1.60)

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

ChIP-seq reads were aligned to the hg19 genome assembly using bwa (version 0.6.1-r104) under default parameters for most experiments except for calibrated ChIP-seq in depleted cells and Re-ChIP, for which bowtie2 (version 2.3.3.1). Unmapped reads were removed using samtools (version 1.3.1) running 'samtools view -F 4'. Reads were sorted and replicates removed using picardtools (version 1.60). Peak calling was performed using macs2 (version 2.1.1.20160309) setting following parameters: '-q 0.05' and '-extsize (value obtained from macs2 predictd step)' and using input as the control.