

# Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization

Aleksandar Kojic<sup>1,7</sup>, Ana Cuadrado<sup>1,7\*</sup>, Magali De Koninck<sup>1</sup>, Daniel Giménez-Llorente<sup>1</sup>,  
Miriam Rodríguez-Corsino<sup>1</sup>, Gonzalo Gómez-López<sup>2</sup>, François Le Dily<sup>3,4</sup>, Marc A. Marti-Renom<sup>3,4,5,6\*</sup>  
and Ana Losada<sup>1\*</sup>

**Two variant cohesin complexes containing SMC1, SMC3, RAD21 and either SA1 (also known as STAG1) or SA2 (also known as STAG2) are present in all cell types. We report here their genomic distribution and specific contributions to genome organization in human cells. Although both variants are found at CCCTC-binding factor (CTCF) sites, a distinct population of the SA2-containing cohesin complexes (hereafter referred to as cohesin-SA2) localize to enhancers lacking CTCF, are linked to tissue-specific transcription and cannot be replaced by the SA1-containing cohesin complex (cohesin-SA1) when SA2 is absent, a condition that has been observed in several tumors. Downregulation of each of these variants has different consequences for gene expression and genome architecture. Our results suggest that cohesin-SA1 preferentially contributes to the stabilization of topologically associating domain boundaries together with CTCF, whereas cohesin-SA2 promotes cell-type-specific contacts between enhancers and promoters independently of CTCF. Loss of cohesin-SA2 rewires local chromatin contacts and alters gene expression. These findings provide insights into how cohesin mediates chromosome folding and establish a novel framework to address the consequences of mutations in cohesin genes in cancer.**

In addition to mediating sister chromatid cohesion, cohesin contributes to the spatial organization of the genome in chromatin loops and topologically associating domains (TADs)<sup>1–5</sup>. In vertebrate somatic cells, cohesin complexes carry one of two versions of the SA subunit, namely SA1 or SA2, which are encoded by *Stag1* and *Stag2*, respectively<sup>6</sup>. Studies in human and mouse cells indicate that cohesin-SA1 and cohesin-SA2 are specifically required for telomere and centromere cohesion, respectively<sup>7,8</sup>. Nevertheless, the cohesion provided by either variant complex is sufficient to allow cell proliferation<sup>9</sup>. Whether the two variants have specific roles in genome organization and gene regulation is unclear, although mouse *Stag1*-null embryos die before birth and show altered cohesin distribution and gene expression<sup>10</sup>. Notably, loss-of-function mutations in the *STAG2* gene have been identified in several human cancers, including bladder cancer, Ewing sarcoma and acute myeloid leukemia<sup>11</sup>. Although cohesin-SA1 is sufficient to perform the essential functions of cohesin in *STAG2*-deficient cancer cells<sup>9</sup>, it may not be able to compensate for other non-essential cohesin-SA2 functions. Consistent with this idea, current evidence suggests that the contribution of cohesin dysfunction to tumorigenesis is not related to cohesion defects or genome instability<sup>12–14</sup>, but rather to altered gene regulation<sup>15,16</sup>.

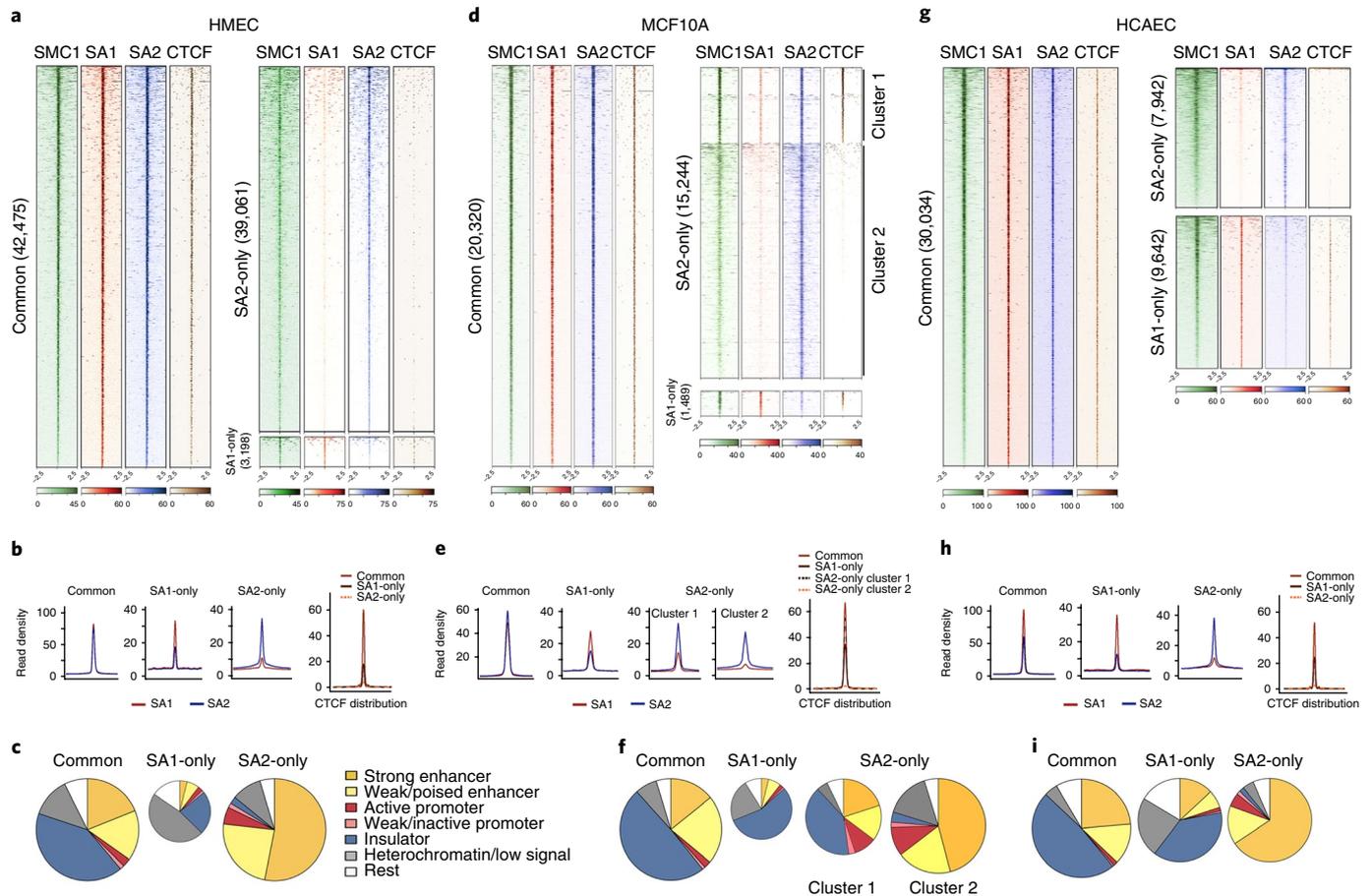
How cohesin affects gene expression remains poorly understood. Analysis of cohesin distribution in mammalian cells shows a large overlap with the sites occupied by the architectural protein CTCF<sup>17–19</sup>. Cohesin and CTCF are present at the boundaries of TADs, sub-megabase regions identified in whole-genome chromatin-conformation capture (Hi-C) experiments that encompass DNA sequences interacting more frequently with sequences inside

than outside the domain<sup>5,20</sup>. TADs are thought to regulate transcription by facilitating interactions between enhancers and promoters present in the same TAD while preventing interactions between elements from different TADs. Deletion of CTCF sites at TAD boundaries changes local topology and affects gene expression<sup>21,22</sup>. A model for TAD generation proposes that, after loading, cohesin extrudes DNA to generate progressively longer chromatid loops until it dissociates from chromatin by the action of cohesin release factor WAPL or until it reaches an obstacle, such as CTCF bound to chromatin, where it gets stalled<sup>23–26</sup>. Cohesin and CTCF are also found inside TADs and contribute to cell-type-specific sub-TAD organization<sup>3</sup>. Moreover, cohesin non-CTCF sites have also been identified in which the complex occupies regions bound by tissue-specific transcription factors or transcriptional regulators such as Mediator<sup>27–29</sup>. In most of these studies, the potential differences between the two variant cohesin complexes were not addressed. We therefore set out to analyze the distribution of cohesin-SA1 and cohesin-SA2 in nontumor human cells, as well as the consequences of their specific downregulation in gene expression and chromatin architecture. Our results reveal important differences between the behavior of the two complexes in the sites they occupy, the dynamics of their chromatin association, their interaction partners and, as a consequence, their contribution to 3D genome organization.

## Results

**Cohesin non-CTCF sites carry SA2 and are present at enhancers.** To characterize the specific roles of cohesin-SA1 and cohesin-SA2 in chromatin architecture, we selected a primary cell line with comparable levels of the two variant complexes, human mammary

<sup>1</sup>Chromosome Dynamics Group, Molecular Oncology Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. <sup>2</sup>Bioinformatics Unit, Structural Biology and Biocomputing Program, Spanish National Cancer Research Centre (CNIO), Madrid, Spain. <sup>3</sup>Centre de Regulació Genòmica (CRG), Barcelona Institute for Science and Technology (BIST), Barcelona, Spain. <sup>4</sup>Universitat Pompeu Fabra, Barcelona, Spain. <sup>5</sup>CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. <sup>6</sup>ICREA, Barcelona, Spain. <sup>7</sup>These authors contributed equally: Aleksandar Kojic, Ana Cuadrado. \*e-mail: [acuadrado@cnio.es](mailto:acuadrado@cnio.es); [martirenom@cnag.crg.eu](mailto:martirenom@cnag.crg.eu); [alosada@cnio.es](mailto:alosada@cnio.es)

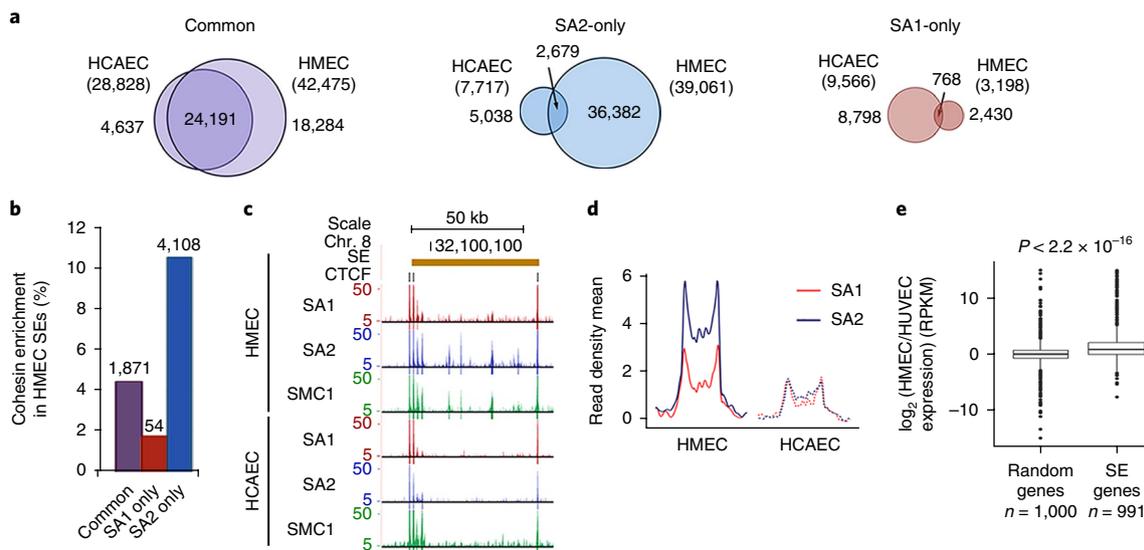


**Fig. 1 | A large fraction of cohesin-SA2 localizes to enhancers independently of CTCF. a**, Analysis of ChIP-seq read distribution for SA1, SA2, SMC1 and CTCF around common, cohesin-SA1-only and cohesin-SA2-only positions within a 5-kb window in HMECs. Color bars below heat maps indicate ChIP-seq read number. **b**, Average read density plots for SA1 (red) and SA2 (blue) distribution in common, SA1-only and SA2-only positions, as well as for CTCF. **c**, Pie charts showing the distribution of cohesin positions in chromatin states, as defined in HMECs. **d–f**, ChIP-seq read distribution analyses (**d**), average read density plots (**e**) and distribution of cohesin positions in chromatin states (**f**) as in **a–c**, respectively, but in MCF10A cells. **g–i**, ChIP-seq read distribution analyses (**g**), average read density plots (**h**) and distribution of cohesin positions in chromatin states (**i**) as in **a–c**, respectively, but in HCAECs. The CTCF datasets are from ENCODE (Supplementary Dataset 2).

epithelial cells (HMECs; Supplementary Fig. 1). We first analyzed the genomic distribution of SMC1, SA1 and SA2 by chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) using custom-made, validated antibodies and high-depth sequencing (about 100 million reads) to ensure whole-genome coverage (Supplementary Dataset 2). Reads were aligned to the reference genome, and peaks were called using MACS2 (false discovery rate (FDR) < 0.01). Overlaps between the peaks obtained in the SA1-specific and SA2-specific immunoprecipitates defined three categories: common, SA2-only and SA1-only positions. Common cohesin positions (42,475) were occupied by either variant complex and colocalized with CTCF (Fig. 1a). They were featured by high cohesin occupancy and similar read density for SA1 and SA2 (Fig. 1b). In contrast, most of the SA2-only cohesin positions (39,061) had no or very little CTCF and a lower read density. The fraction of SA1-only positions was small (3,198) and contained some SA2 and CTCF (Fig. 1a (lower right), b). Analysis of the distribution of these cohesin-binding sites in chromatin states defined by ChromHMM in HMECs<sup>30</sup> revealed that most of the SA2-only cohesin positions (77%) were in enhancers, particularly in active ones (Fig. 1c). The distribution of the common positions was very different, with only 35% present in enhancers, whereas another 41% were in insulators defined by the sole presence of CTCF. Some SA1-only

positions were in insulators (23%) and enhancers (10%), but most were present in a chromatin state that was designated as ‘heterochromatin, low signal’<sup>30</sup>. Motif discovery analysis showed that both common and SA1-only positions were significantly enriched for the CTCF-binding motif, whereas SA2-only positions were populated by recognition motifs of several transcription factors other than CTCF (Supplementary Fig. 2a).

We validated the findings above in MCF10A cells, a nontumorigenic epithelial breast cell line that, unlike HMECs, can be easily grown and transfected for functional analyses. Common positions had similar average read densities for SA1 and SA2 and overlapped with CTCF (Fig. 1d,e). Among the SA2-only positions assigned by peak calling, read distribution heat maps distinguished two clusters (Fig. 1d). Although the cohesin positions in both clusters were enriched in SA2, those in cluster 1 contained some SA1 and CTCF (Fig. 1d), and its distribution among chromatin states was not very different from that of the common and SA1-only positions (Fig. 1f). The larger cluster 2, in contrast, grouped true SA2-only positions—as in HMECs, these positions lacked CTCF and were enriched in enhancers and depleted in insulators as compared with the common and SA1-only positions (Fig. 1e,f). Cohesin-SA2 may have partners other than CTCF at enhancers and promoters, most likely transcription factors. Consistent with this possibility, proteomic



**Fig. 2 | Cohesin-SA2-only positions are enriched in cell-type-specific super-enhancers.** **a**, Venn diagrams showing overlap of cohesin-binding sites between HMECs and HCAECs. Common positions are more conserved. **b**, Cohesin enrichment in super-enhancers (SEs) defined in HMECs. **c**, Example of cohesin distribution in HMEC and HCAEC cells within a HMEC-specific super-enhancer. **d**, Plot showing SA1 and SA2 enrichment in HMEC and HCAEC cells along all HMEC super-enhancers. **e**, Box plot comparing changes in expression between random genes and genes associated with HMEC-specific super-enhancers<sup>32</sup>. Boxes represent interquartile range (IQR); the midline represents the median; whiskers are  $1.5 \times$  IQR; and individual points are outliers. Statistical significance was calculated with a Wilcoxon signed-rank test.

analyses of immunoprecipitates obtained from MCF10A cell extracts with anti-SA1 and anti-SA2 identified several transcriptional regulators that interacted with SA2 and not SA1, including ZMYM2 and YAP1 (Supplementary Dataset 3). ZMYM2 acts as a co-repressor in association with the LSD1-CoREST-HDAC1/2 complex, whereas YAP1 is a co-activator. ChIP-seq analyses for ZMYM2 (this study) and activated YAP1<sup>31</sup> confirmed the presence of SA2, and not SA1 or CTCF, at their binding sites in MCF10A cells (Supplementary Fig. 2b). We conclude that cohesin can be found at CTCF sites and non-CTCF sites, and that in the latter case, cohesin-SA2 is the predominant variant. These cohesin-SA2 non-CTCF positions are enriched in cis-regulatory elements co-occupied by transcriptional regulators.

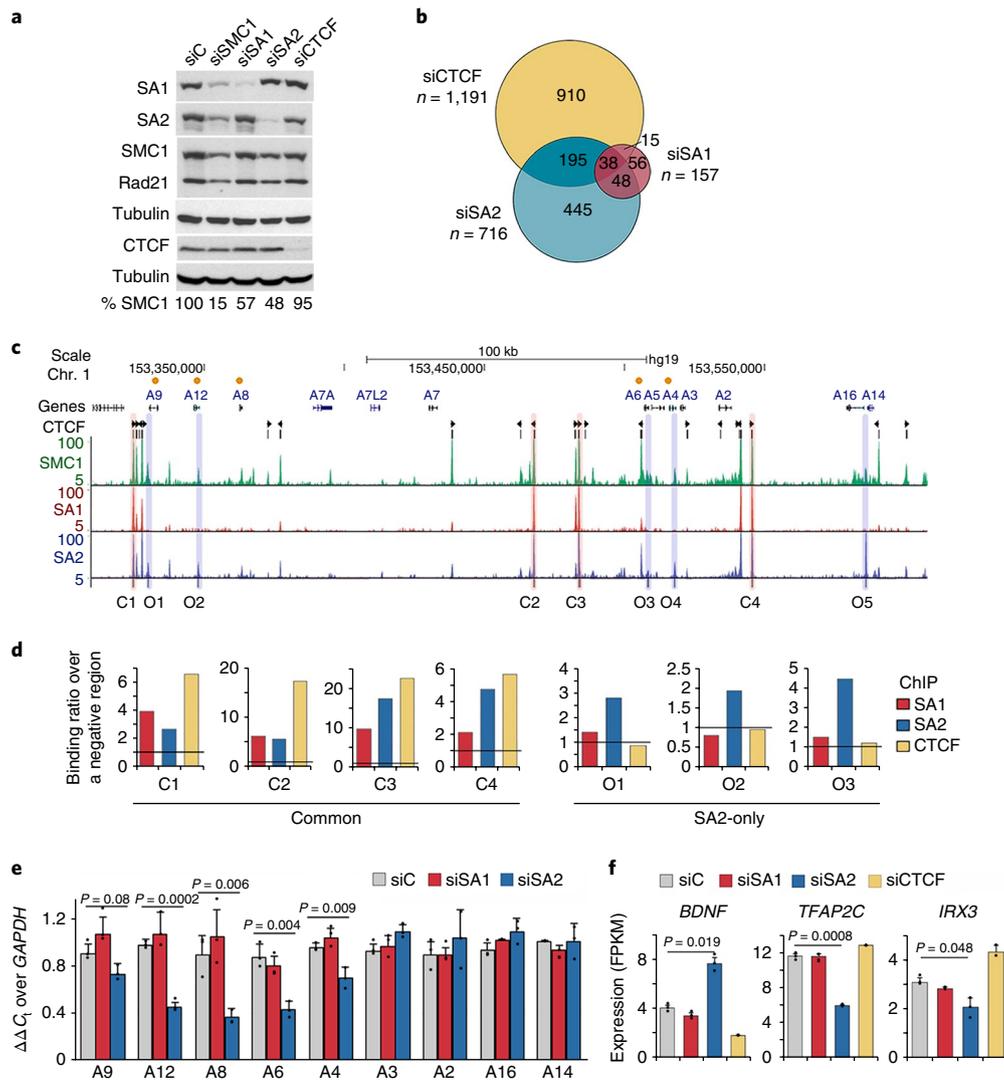
**Cohesin-SA2 is linked to tissue-specific transcription.** We determined the distribution of cohesin-SA1 and cohesin-SA2 in a third cell line that was of different embryonic origin, human cardiac endothelial cells (HCAECs). Here the number of SA1-only and SA2-only positions was similar (Fig. 1g). Read density profile plots for SA1 and SA2 in common positions suggested that ChIP with anti-SA2 had been less efficient in these cells (Fig. 1h). We suspect that this has two consequences: (i) SA2-only positions with low cohesin occupancy go undetected, and (ii) a fraction of the positions assigned as SA1 only by peak calling are in fact common positions. In any case, as in the other two cell lines, common and most SA1-only positions overlapped with CTCF, whereas SA2-only positions lacked CTCF. The distribution of SA1-only positions among chromatin states was close to that of the common positions, with a prevalence in insulators, whereas SA2-only positions were enriched in enhancers, as previously described for epithelial cells (Fig. 1i).

We observed that a large fraction of common positions was conserved between the epithelial and endothelial cells, whereas SA1-only and SA2-only positions were not (Fig. 2a). Moreover, cohesin-SA2-only sites in HMECs were particularly enriched in super-enhancers defined in the same cell line, which control genes associated with cell identity<sup>32</sup> (Fig. 2b,c). SA2 signals were enriched relative to SA1 signals in active super-enhancers (Fig. 2d), and

loss of SA2 at these super-enhancers in HCAECs correlated with decreased expression of their associated genes (Fig. 2e).

To further understand the effect of each cohesin variant on gene regulation, we transfected MCF10A cells with siRNAs that targeted the transcripts of the genes encoding SA1 (siSA1) and SA2 (siSA2) and, for comparison, CTCF (siCTCF) and SMC1 (siSMC1). Comparable depletion of SA1 or SA2 left similar amounts of cohesin (SMC1) in the cells (Fig. 3a). By using a stringent criterion for RNA sequencing (RNA-seq) data analysis, we identified 157 and 716 differentially expressed genes (DEGs) in cells that were treated with siSA1 and siSA2, respectively (Fig. 3b and Supplementary Datasets 4–6). Of the 630 genes that were deregulated only after SA2 depletion, 445 were not affected by knockdown of CTCF expression, which confirms a CTCF-independent role for SA2 in the control of gene expression. Among the genes that were deregulated in siSA2-treated cells, there were several encoding members of the S100 family of calcium-binding proteins, which are located in a 300-kb-long gene cluster on chromosome 1 (Fig. 3c (orange dots) and Supplementary Dataset 5). This region contains strong common cohesin peaks, as well as less-prominent cohesin-SA2-only binding sites at the promoters of the deregulated genes (Fig. 3c). We used this locus to validate the ChIP-seq data by ChIP-qPCR (Fig. 3d) and the RNA-seq data by qRT-PCR (Fig. 3e). Other genes whose expression was affected by SA2 downregulation were brain-derived neurotrophic factor (*BDNF*), a known target of CoREST in non-neuronal cells<sup>33</sup>, and those encoding two of the top ten core transcription factors proposed to control cell identity in mammary gland cells<sup>34</sup>, IRX3 and TFAP2C (Fig. 3f). Gene set enrichment analyses also revealed aberrant upregulation of pathways specific to the hematopoietic system and the nervous system in MCF10A cells after siSA2 treatment (Supplementary Fig. 3). Taken together with the preferential enrichment of cohesin-SA2 at super-enhancers, these pieces of evidence support a contribution of cohesin-SA2 to tissue-specific gene expression.

**Different dynamic behavior of cohesin-SA1 and cohesin-SA2.** ChIP-seq read-density plots of SMC1 distribution around common

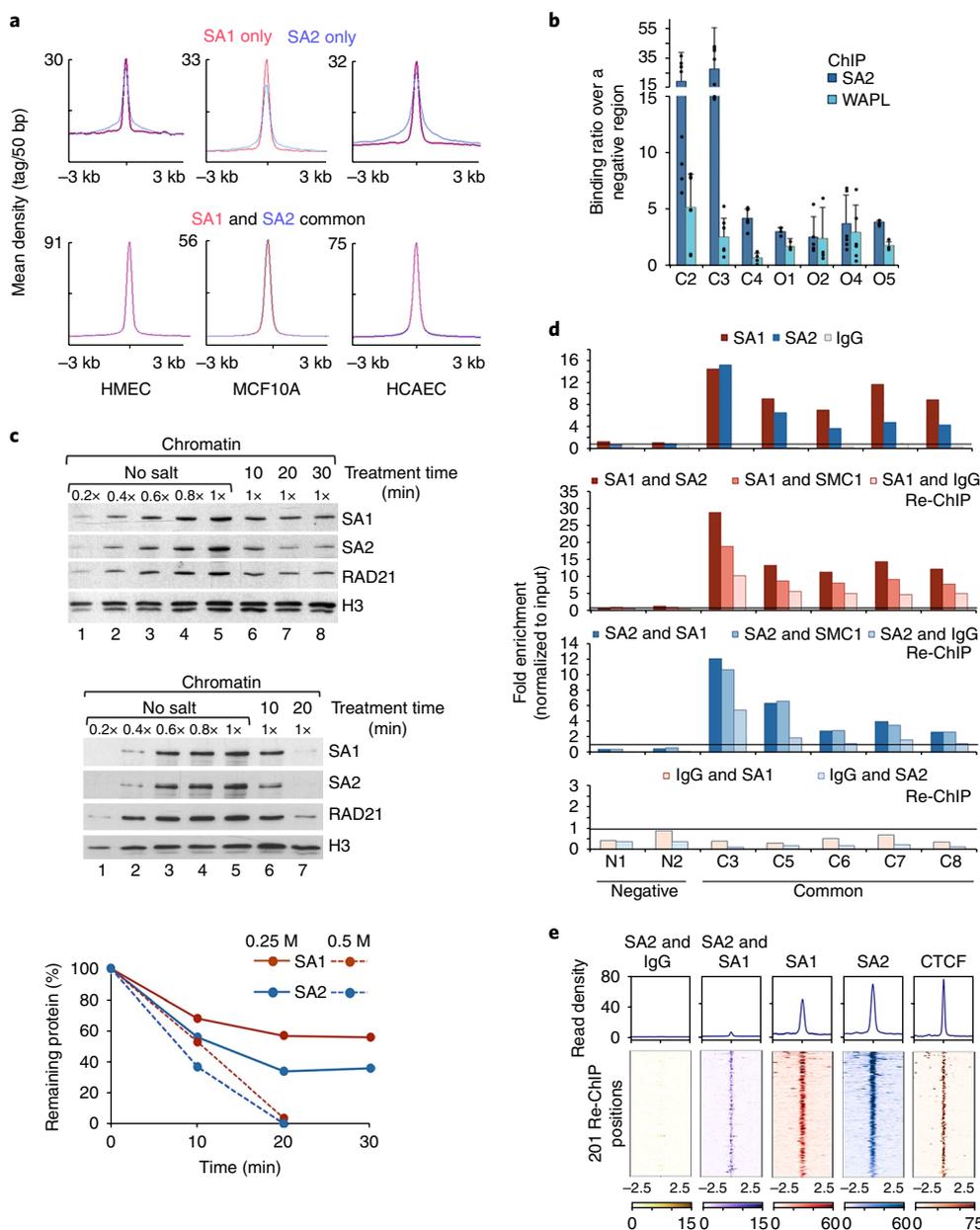


**Fig. 3 | SA2-specific changes in transcription are related to cell identity.** **a**, Representative immunoblot showing levels of cohesin and CTCF after transfection of the indicated siRNAs in MCF10A cells (uncropped blot images are shown in Supplementary Dataset 1).  $n = 3$  experiments. **b**, Venn diagram showing the overlap between genes deregulated after downregulation of SA1, SA2 or CTCF, relative to mock-transfected cells (FDR < 0.05;  $\log_2$  (fold change) < -0.5 or  $\log_2$  (fold change) > 0.5; FPKM > 3; in at least one condition). **c**, UCSC browser image of the *S100A* gene cluster showing genes (orange dots indicate those deregulated in SA2-depleted cells), CTCF peaks (and motif orientation), and genomic distribution of SMC1, SA1 and SA2 in MCF10A cells. Positions corresponding to common (C) and SA2-only (O) cohesin-binding sites used in ChIP-qPCR analyses are shadowed in red and blue, respectively. **d**, ChIP-qPCR validation of SA1, SA2 and CTCF binding to common and SA2-only positions. Bars represent the mean of three technical replicates. **e, f**, Gene expression levels of *S100A* genes assessed by RT-qPCR and normalized to levels of the housekeeping gene *GAPDH* (**e**) and those encoding cell-type-specific transcription factors assessed by RNA-seq (**f**) in control cells and cells transfected with siSA1, siSA2 or siCTCF. In **e** and **f**, data are means and s.d. from three independent experiments. A Student's *t* test was used to assess statistical significance. FPKM, fragments per kilobase of transcript per million mapped reads.

and cohesin-SA1-only positions produced sharp and narrow profiles in all of the three cell lines analyzed, whereas for SA2-only positions the profiles were broader (Fig. 4a). These data suggest that the cohesin-SA2 present at these positions is more dynamic. Consistent with this possibility, quantitative ChIP-qPCR analyses showed that cohesin-SA2 at common positions were less likely to associate with WAPL, a factor that dissociates cohesin from chromatin<sup>35</sup>, as compared to those present at the SA2-only positions (Fig. 4b). Moreover, WAPL removal in HAP1 cells<sup>23</sup> increased SMC1 occupancy more in cohesin non-CTCF sites, which were most likely bound by cohesin-SA2, than in cohesin CTCF sites (Supplementary Fig. 4). There was also more WAPL in anti-SA2 immunoprecipitates than in anti-SA1 immunoprecipitates (highlighted in Supplementary Dataset 3).

To further test our hypothesis that cohesin-SA2 is more dynamic than cohesin-SA1, we performed a salt-extraction experiment. The chromatin fraction of MCF10A cells was treated with 0.25 M or 0.5 M NaCl for 10 or 20 min, and the amount of each variant that remained on the chromatin was assessed by immunoblotting. We found that SA2 was more sensitive to salt than SA1, as seen at all time points in the treatment with the lower salt concentration (Fig. 4c, top and bottom). After treatment with the higher salt concentration, the enhanced sensitivity of SA2 could be seen at the earlier time point (Fig. 4c, middle and bottom). We conclude that the association of cohesin-SA2 with chromatin is less tight, or more dynamic, than the association of cohesin-SA1.

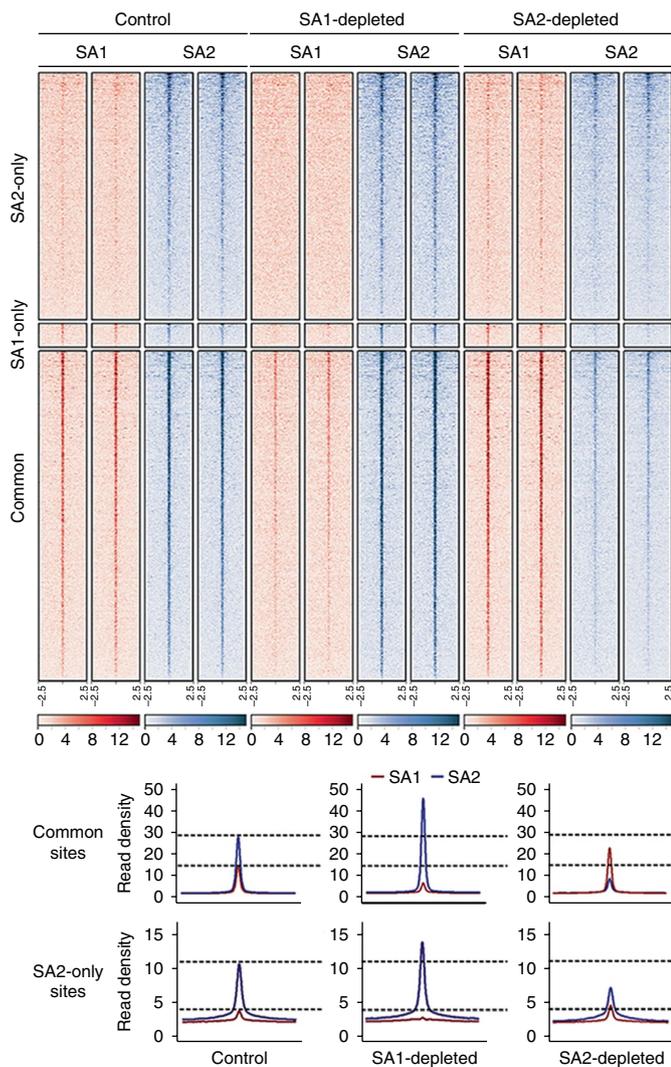
Both cohesin-SA1 and cohesin-SA2 can be found at common cohesin-binding sites. This may be because cells in a population can



**Fig. 4 | Different behavior of cohesin in common and SA2-only positions.** **a**, Cohesin SMC1 distribution in HMECs, MCF10A cells and HCAECs. Maximum mean tag density is indicated on the y axis. **b**, SA2 and WAPL binding to the indicated cohesin positions from the *S100* locus. Data represent the mean and s.d. of at least three independent experiments performed in triplicate. **c**, Chromatin-bound cohesin (SA1, SA2 and the common cohesin subunit RAD21) was determined after treatment with 0.25 M NaCl (top) or 0.5 M NaCl (middle) at different time points. Quantification for SA1 and SA2 signal normalized to histone H3 signal is shown in the graph (bottom). Uncropped blot images are shown in Supplementary Dataset 1. **d,e**, The simultaneous presence of at least two cohesin complexes at a given position within the same cell was assayed by Re-ChIP-qPCR (**d**) and confirmed by Re-ChIP-seq (**e**). A single experiment was performed. Chromatin eluted from the first ChIP with anti-SA1 or anti-SA2 was incubated with anti-SA2 or anti-SA1, respectively, as well as with anti-SMC1 or IgG as positive and negative controls, respectively. In **d**, the bottom graph shows Re-ChIP of chromatin eluted from IgG beads with anti-SA1 and anti-SA2. Positions C3–C8 are ‘common’ cohesin-binding sites; N1 and N2 are negative regions. All of the positions captured by Re-ChIP-seq in **e** correspond to common sites in MCF10A cells. Color bars indicate ChIP-seq read number.

have either variant complex or because both complexes can coexist at a given position within a cell. Re-ChIP (also known as sequential ChIP) experiments with anti-SA1 and anti-SA2 revealed that at least two independent cohesin rings can coexist in the same genomic position in the same cell (Fig. 4d,e). We speculate that stacking at CTCF-bound sites may contribute to stabilize cohesin binding by preventing access of WAPL to cohesin. Alternatively, CTCF itself may stop cohesin progression<sup>36</sup> and at the same time prevent its dissociation by an as yet unclear mechanism.

**Cohesin-SA1 cannot occupy SA2-only sites.** Next, we asked how cohesin distribution changed after depletion of SA1 or SA2. Calibrated ChIP-seq analyses with anti-SA1 and anti-SA2 were performed in cells that were mock-depleted or depleted of SA1 or SA2. In SA1-depleted cells, there was little cohesin-SA1 left at any position, whereas the presence of cohesin-SA2 increased both at common and SA2-only sites as compared to that in mock-transfected cells, and even at SA1-only sites (Fig. 5). It is likely that these SA1-only sites, defined based on peak calling (Fig. 1), are in fact common



**Fig. 5 | Cohesin-SA1 cannot occupy SA2-only sites.** Read heat map (top) and read density plots (bottom) showing SA1 and SA2 distribution around the cohesin positions defined in Fig. 1d in control cells and in cells that were treated with siSA1 or siSA2. Two independent replicates were performed for each condition. Read density plots were built by merging the reads from the two replicates.

positions in which SA2 is immunoprecipitated less efficiently. Of note, cohesin-SA1 could not occupy cohesin-SA2-only sites in SA2-depleted cells, and instead it accumulated further at common positions. We conclude that both cohesin-SA1 and cohesin-SA2 can reach common binding sites independently of each other. Thus, in the absence of one variant, the other could, in principle, compensate for its loss at these CTCF-bound sites. In contrast, cohesin-SA1 cannot occupy cohesin-SA2-only positions when SA2 is missing.

**Cohesin-SA1 and cohesin-SA2 make different contributions to genome architecture.** To address the consequences of SA1 or SA2 depletion on genome architecture, we performed Hi-C experiments in MCF10A cells that were depleted for SA1 or SA2 (Fig. 6a, Supplementary Fig. 5a,b and Supplementary Dataset 7). The identity of active (A) and repressive (B) compartments<sup>37</sup> was mostly preserved (Fig. 6b and Supplementary Fig. 5c). TAD number increased in 204 TADs after SA1 depletion but decreased in 439 TADs after SA2 depletion (Fig. 6c). TAD border strength was decreased, particularly

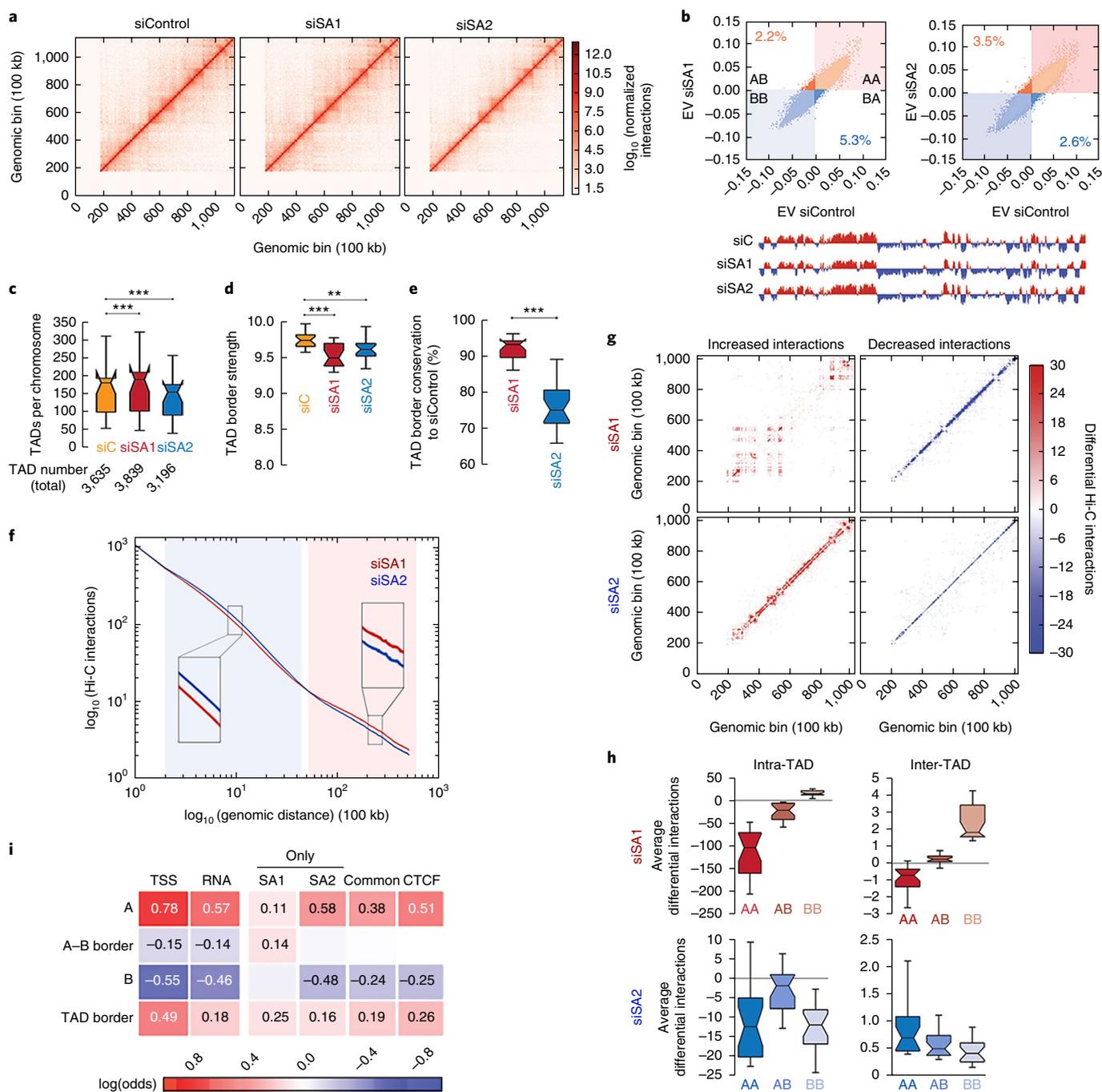
in SA1-depleted cells (Fig. 6d), whereas TAD border conservation was diminished by 25% after SA2 depletion (Fig. 6e and Supplementary Fig. 5d). We therefore suggest that some TADs might arise in a CTCF-independent manner and instead depend on the interaction of cohesin-SA2 with different transcriptional regulators. Although we could not test this idea with the current resolution of our Hi-C analyses, it agrees with recent data showing that ~20% of TAD borders are maintained after acute elimination of CTCF in mouse embryonic stem (ES) cells<sup>38</sup>, as well as with high-resolution Hi-C maps from the same cells revealing a set of TAD boundaries featured by the presence of cohesin and active marks but no CTCF<sup>39</sup>.

Analysis of genomic interactions as a function of genomic distance further provided evidence for specific contributions of the two cohesin variants to chromatin architecture (Fig. 6f and Supplementary Fig. 5e). Loss of SA2 increased mid-range contacts (0.1–1.3 Mb), whereas loss of cohesin-SA1 increased long-range contacts (>1.4 Mb). These distinct effects were also evident in matrices that represented separately gained and lost interactions for each condition relative to control cells (Fig. 6g and Supplementary Fig. 6). SA1 depletion increased very long-range interactions, most of which were located within the B compartment, whereas mid-range interactions within the A compartment were lost (Fig. 6g,h (top)). One possible interpretation of these data is that SA1 depletion results in a more ‘relaxed’ A compartment, which is compensated by increased compaction of the B compartment. In contrast, SA2 depletion increased inter-TAD mid-range contacts, mostly within the A compartment (Fig. 6g,h (bottom)), at least in part owing to loss of TAD borders. SA2 depletion also decreased short-range intra-TAD contacts, which could correspond to enhancer-enhancer or enhancer-promoter interactions, given the prevalence of SA2-only positions in these elements. Finally, the specific enrichment of cohesin-SA1-only positions in A–B borders (Fig. 6i) prompted us to speculate that cohesin-SA1 might have a unique role in modulating A–B compartment identity. However, visual examination of the Hi-C matrices and the resulting eigenvalues used in compartment analyses did not reveal compartment switches in any of the conditions (Fig. 6b, lower part).

To interpret our results, we propose that cohesin-SA1 has a more structural role in genome organization, by supporting TAD or sub-TAD formation together with CTCF, whereas cohesin-SA2 is more critical for functional intra-TAD contacts together with transcriptional regulators. In the absence of cohesin-SA1, cohesin-SA2 can still cooperate with CTCF in genome organization, although border strength is decreased and the A compartment is loosened. In the absence of cohesin-SA2, short-range intra-TAD contacts decrease, whereas new contacts are formed between neighboring TADs, and these changes have more noticeable consequences for gene expression.

## Discussion

Recent studies in different cellular systems in which cohesin or CTCF depletion was performed have led to the conclusion that TADs and compartments arise independently<sup>38,40–42</sup>. TADs would depend on cohesin and CTCF, whereas genomic compartmentalization would rely mostly on epigenetic features regardless of chromatin contacts. Although TAD boundaries are largely invariant across cell types<sup>5</sup>, the specific interactions within TADs may not be<sup>43</sup>. Moreover, results from single-cell Hi-C experiments imply a certain degree of stochasticity in the TAD boundary definition among cells in the population<sup>44,45</sup>. To our knowledge, our results show for the first time that the two variant cohesin complexes have nonredundant functions in genome organization. After downregulation of one or the other, the changes that we observed were, not unexpectedly, different from those observed after removal of all cohesin<sup>40–42</sup>. The amount of total cohesin present on chromatin in siSA1-treated and siSA2-treated cells was very similar, whereas the



**Fig. 6 | Distinct contribution of cohesin-SA1 and cohesin-SA2 to genome architecture.** **a**, Vanilla-normalized Hi-C matrices for chromosome 13 at 100-kb resolution in MCF10A cells. All analyses in this figure use data merged from two replicates. Separate analyses for each replicate are shown in Supplementary Figs. 5 and 6. **b**, Scatterplot of eigenvectors (EV) of the intrachromosomal interaction matrices indicated. Numbers within the plot show the percentage of bins that changed compartment. The first eigenvector for chromosome 13 at 100-kb resolution is shown below the plots. **c–e**, Box plots showing the number of TADs per chromosome (**c**), TAD border strength (**d**) and TAD border conservation (**e**). The box extends from the lower to upper quartile values of the data, with a line at the median. Notches represent the confidence interval around the median. Box plots are for the chromosome average values ( $n = 23$ ). **f**, Hi-C interactions as a function of genomic distance averaged across the genome for a maximum distance of 50 Mb. **g**, Matrices showing increased (red) and decreased (blue) interactions in chromosome 15 when comparing siSA1-treated cells (top) or siSA2-treated cells (bottom) to control cells. **h**, Effect of the SA1 or SA2 depletion in differential inter-TAD and intra-TAD interactions in different compartments. Box plots are for the chromosome average values ( $n = 23$ ). **i**, Enrichment of SA1-only, SA2-only and common sites in A/B compartments, A-B borders and TAD borders. Squares with numbers are significant ( $P < 0.001$  by Fisher’s exact test).

relative abundance of each variant changed dramatically, bringing about the changes in cohesin distribution, chromatin contacts and gene expression reported above.

Previous analyses have shown that cohesin colocalizes with transcription factors independently of CTCF and thereby contributes to tissue-specific transcription<sup>29</sup>. Here we show that cohesin-SA2 is the

prevalent variant at cohesin non-CTCF sites and confirm that these SA2-only sites tend to be tissue specific and are enriched at enhancers and super-enhancers. Notably, cohesin-SA1 cannot replace cohesin-SA2 at non-CTCF sites. The mechanisms that position SA1- and SA2-containing complexes remain to be identified. The two SA subunits are highly similar, with over 70% sequence identity along the central part of the protein. The homology decreases in the N- and C-terminal regions, and, for instance, SA1 but not SA2 interacts with the telomeric protein TRF1 through its N terminus<sup>46</sup>. In the same way, SA2 may interact with certain transcriptional regulators through its unique regions. Alternatively, chromatin loops between enhancers and promoters and between CTCF sites may arise by distinct mechanisms, the latter being possibly loop extrusion, and the two SA subunits may be preferentially used for one or the other. In this regard, it is worth mentioning recent *in vitro* data that show that establishment of DNA-DNA interactions by a cohesin ring already embracing double-stranded DNA requires the second DNA molecule to be single-stranded DNA<sup>47</sup> and that purified SA2 binds single-stranded DNA better than SA1<sup>48</sup>. One could envision cohesin-SA2 interacting with enhancer RNA (eRNA) to stabilize an enhancer-promoter loop<sup>49</sup>.

Somatic mutations in *STAG2* have been reported in multiple human cancers, most prominently bladder, Ewing sarcoma and myeloid malignancies<sup>11</sup>. The presence of cohesin-SA1 allows *STAG2*-deficient cancer cells to survive by ensuring sufficient cohesion between the sister chromatids<sup>9</sup>. However, cohesin-SA1 cannot occupy SA2-only sites involved in enhancer-promoter interactions, and, as a consequence, expression of some key genes may be altered. Recent studies have shown that elimination of all chromatin loops mediated by cohesin has little effect on steady-state transcription<sup>40,42</sup>; yet, cohesin may be most relevant for transcriptional responses induced after differentiation or lineage commitment<sup>50</sup>, and when deregulated, it may contribute to tumorigenesis, as shown for hematopoietic stem and progenitor cells<sup>51</sup>.

## Methods

Methods, including statements of data availability and any associated accession codes and references, are available at <https://doi.org/10.1038/s41594-018-0070-4>.

Received: 21 August 2017; Accepted: 23 April 2018;

Published online: 04 June 2018

## References

- Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43**, 525–558 (2009).
- Hadjur, S. et al. Cohesins form chromosomal cis-interactions at the developmentally regulated *IFNG* locus. *Nature* **460**, 410–413 (2009).
- Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
- Zuin, J. et al. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* **111**, 996–1001 (2014).
- Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- Losada, A., Yokochi, T., Kobayashi, R. & Hirano, T. Identification and characterization of SA or Scc3p subunits in the *Xenopus* and human cohesin complexes. *J. Cell Biol.* **150**, 405–416 (2000).
- Canudas, S. & Smith, S. Differential regulation of telomere and centromere cohesion by the Scc3 homologs SA1 and SA2, respectively, in human cells. *J. Cell Biol.* **187**, 165–173 (2009).
- Remeseiro, S. et al. Cohesin-SA1 deficiency drives aneuploidy and tumorigenesis in mice due to impaired replication of telomeres. *EMBO J.* **31**, 2076–2089 (2012).
- van der Lelij, P. et al. Synthetic lethality between the cohesin subunits *STAG1* and *STAG2* in diverse cancer contexts. *eLife* **6**, e26980 (2017).
- Cuadrado, A., Remeseiro, S., Gómez-López, G., Pisano, D. G. & Losada, A. The specific contributions of cohesin-SA1 to cohesion and gene expression: implications for cancer and development. *Cell Cycle* **11**, 2233–2238 (2012).
- De Koninck, M. & Losada, A. Cohesin mutations in cancer. *Cold Spring Harb. Perspect. Med.* **6**, a026476 (2016).
- Balbás-Martínez, C. et al. Recurrent inactivation of *STAG2* in bladder cancer is not associated with aneuploidy. *Nat. Genet.* **45**, 1464–1469 (2013).
- Solomon, D. A. et al. Frequent truncating mutations of *STAG2* in bladder cancer. *Nat. Genet.* **45**, 1428–1430 (2013).
- Kon, A. et al. Recurrent mutations in multiple components of the cohesin complex in myeloid neoplasms. *Nat. Genet.* **45**, 1232–1237 (2013).
- Viny, A. D. et al. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J. Exp. Med.* **212**, 1819–1832 (2015).
- Corces, M. R. & Corces, V. G. The three-dimensional cancer genome. *Curr. Opin. Genet. Dev.* **36**, 1–7 (2016).
- Wendt, K. S. et al. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796–801 (2008).
- Parelho, V. et al. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422–433 (2008).
- Rubio, E. D. et al. CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA* **105**, 8309–8314 (2008).
- Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation center. *Nature* **485**, 381–385 (2012).
- Guo, Y. et al. CRISPR inversion of CTCF sites alters genome topology and enhancer-promoter function. *Cell* **162**, 900–910 (2015).
- Narendra, V. et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
- Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707 (2017).
- Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456–E6465 (2015).
- Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
- de Wit, E. et al. CTCF binding polarity determines chromatin looping. *Mol. Cell* **60**, 676–684 (2015).
- Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
- Schmidt, D. et al. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20**, 578–588 (2010).
- Faure, A. J. et al. Cohesin regulates tissue-specific expression by stabilizing highly occupied cis-regulatory modules. *Genome Res.* **22**, 2163–2175 (2012).
- Ernst, J. et al. Mapping and analysis of chromatin-state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
- Jang, W., Kim, T., Koo, J. S., Kim, S. K. & Lim, D. S. Mechanical-cue-induced YAP instructs SKP2-dependent cell cycle exit and oncogenic signaling. *EMBO J.* **36**, 2510–2528 (2017).
- Whyte, W. A. et al. Master transcription factors and Mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
- Ballas, N., Grunseich, C., Lu, D. D., Speth, J. C. & Mandel, G. REST and its corepressors mediate plasticity of neuronal gene chromatin throughout neurogenesis. *Cell* **121**, 645–657 (2005).
- D'Alessio, A. C. et al. A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Rep.* **5**, 763–775 (2015).
- Tedeschi, A. et al. WAPL is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564–568 (2013).
- Davidson, I. F. et al. Rapid movement and transcriptional relocation of human cohesin on DNA. *EMBO J.* **35**, 2671–2685 (2016).
- Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
- Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
- Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
- Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
- Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
- Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331–336 (2015).
- Flyamer, I. M. et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544**, 110–114 (2017).
- Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).
- Canudas, S. et al. Protein requirements for sister telomere association in human cells. *EMBO J.* **26**, 4867–4878 (2007).

47. Murayama, Y., Samora, C. P., Kurokawa, Y., Iwasaki, H. & Uhlmann, F. Establishment of DNA–DNA interactions by the cohesin ring. *Cell* **172**, 465–477 (2018).
48. Countryman, P. et al. Cohesin SA2 is a sequence-independent DNA-binding protein that recognizes DNA replication and repair intermediates. *J. Biol. Chem.* **293**, 1054–1069 (2018).
49. Li, W. et al. Functional roles of enhancer RNAs for estrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
50. Ing-Simmons, E. et al. Spatial enhancer clustering and regulation of enhancer-proximal genes by cohesin. *Genome Res.* **25**, 504–513 (2015).
51. Mullenders, J. et al. Cohesin loss alters adult hematopoietic stem cell homeostasis, leading to myeloproliferative neoplasms. *J. Exp. Med.* **212**, 1833–1850 (2015).

### Acknowledgements

We thank Y. Cuartero and J. Quilez (4D Genome–CRG) for technical help with the Hi-C experiments, D. Rico (Newcastle University), F. X. Real (CNIO) and M. Manzanera (CNIC) for comments on the manuscript, T. Hirano (RIKEN) and H. Yu (UT Southwestern) for reagents, and M. Quintela (CNIO) for MCF10A cells. This work has been supported by the Spanish Ministry of Economy and Competitiveness and FEDER funds (grant no. BFU2013-48481-R (A.L.), BFU2016-79841-R (A.L.) and BFU2013-47736-P (M.A.M.-R.), fellowship no. BES-2014-069166 (M.D.K.), and Centro de Excelencia Severo Ochoa grant no. SEV-2015-0510 (to CNIO) and SEV-2012-0208 (to CRG), the European Research Council (FP7/2010–2015, ERC grant agreement

609989; M.A.M.-R.), the EU Horizon 2020 Research and Innovation Program (agreement 676556; M.A.M.-R.), the CERCA Programme–Generalitat de Catalunya (M.A.M.-R.) and the La Caixa Foundation (PhD fellowship to A.K.).

### Author contributions

A.C. and A.K. performed most of the experiments with technical help from M.R.-C.; M.D.K. performed the immunoprecipitation and salt-extraction experiments; E.L.D. and A.K. performed the Hi-C experiments; G.G.-L. analyzed the RNA-seq data; A.K. and D.G.-L. analyzed the ChIP-seq data; M.A.M.-R. analyzed the Hi-C data; A.C. and A.L. planned the project and wrote the manuscript with contributions from all of the authors.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41594-018-0070-4>.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Correspondence and requests for materials** should be addressed to A.C. or M.A.M.-R. or A.L.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Methods

**Cell lines.** Human primary cell lines were purchased from Lonza and cultured according to the manufacturer's recommendations. Normal human astrocytes (NHAs; CC-2565) were grown in ABM basal medium (CC-3187) supplemented with AGM Bulletkit (CC-4123). Skeletal muscle cells (SKMCs; CC-2561) were cultured in SkBM basal medium (CC-3161) supplemented with SkGM Bulletkit (CC-4139). Normal human bronchial epithelial cells (NHBEs; CC-2540) were cultured in BEBM basal medium (CC-3171) supplemented with BEGM Bulletkit (CC-4175). Coronary artery endothelial cells (HCAECs; CC-2585) were grown in EBM2 basal medium (CC-3156) supplemented with EGM2-MV Bulletkit (CC-4147). Normal human epidermal keratinocytes (NHEKs; cat. no. 00192627) were grown in KBM-Gold basal medium (cat. no. 00192151) supplemented with KGM-Gold Bulletkit (cat. no. 00192060). Normal mammary epithelial cells (HMECs; CC-2551) were cultured in MEBM basal medium (CC-3171) supplemented with MEGM Bulletkit (CC-3150). Normal human osteoblasts (NHOst; CC-2538) were grown in OBM basal medium (CC-3208) supplemented with OGM Bulletkit (CC-3207). Prostate epithelial cells (PrECs; CC-2555) were cultured with PrEBM basal medium (CC-3165) supplemented with PrEGM Bulletkit (CC-3166). Human umbilical vein endothelial cells (HUVECs; CC-2517) were grown in EBM basal medium (CC-3121) supplemented with EGM Bulletkit (CC-3124). MCF10A cells (a gift from M. Quintela, CNIO) were grown in DMEM-F12 (cat. no. 31330038, Thermo Fisher) supplemented with 20 ng/ml epidermal growth factor (EGF), 0.5 mg/ml hydrocortisone, 100 ng/ml cholera toxin, 10 mg/ml insulin and 5% horse serum.

**Antibodies.** A rabbit polyclonal antibody recognizing human WAPL was generated by using a recombinant C-terminal fragment of the protein (352 amino acids long), cloned by PCR amplification from full-length WAPL cDNA (a gift from T. Hirano (RIKEN, Japan)). A rat monoclonal antibody was raised against the N-terminal region of mouse SA1 and used for western blotting. Additional custom-made antibodies have been previously described for SA1, SA2 and SMC1<sup>2</sup> and for ZMYM2<sup>23</sup> (a gift from H. Yu (UT Southwestern)). Commercial antibodies used included anti-CTCF (clone 07-729; Millipore), anti-tubulin (clone DM1A; Sigma) and anti-histone-H3 (clone ab1791; Abcam).

**Quantitative immunoblotting in whole-cell extracts and chromatin fractions.** Cells were collected after trypsin treatment, counted, resuspended in SDS-PAGE loading buffer at 10<sup>7</sup> cells/ml, sonicated and boiled. Equal volumes were separated by SDS-PAGE and analyzed by immunoblotting. Chromatin fractionation was performed as described<sup>54</sup>, and fractions were run on SDS gels alongside increasing amounts of recombinant proteins corresponding to C-terminal fragments of human SA1 and SA2, to estimate the amount of each variant subunit<sup>10</sup>. To assess the strength of chromatin association of the cohesin variants, chromatin fractions were treated with modified buffer A (10 mM HEPES, 1.5 mM MgCl<sub>2</sub>, 0.34 M sucrose, 10% glycerol, 1 mM DTT and protease inhibitors) containing 0.25 M or 0.5 M NaCl for 10, 20 or 30 min on ice. Solubilized proteins were separated from insoluble chromatin by low-speed centrifugation (4 min at 1,700g), and the latter was analyzed by immunoblotting.

**Treatment with siRNAs.** MCF10A cells were transfected with 50 nM onTARGETplus SMARTpool siRNAs (Dharmacon L-010638, L-021351, L-006833 and L-020165 to target *STAG1*, *STAG2*, *SMC1* and *CTCF*, respectively) using DharmaFECT reagent 1. Transfection efficiency was first estimated by qRT-PCR 24 h after transfection, and it typically reached more than 90% downregulation (data not shown). Cells were harvested at 72 h, and protein levels were assessed by immunoblotting.

**Chromatin immunoprecipitation followed by sequencing and analysis.** ChIP was performed as previously described<sup>10</sup>, with some modifications. Confluent cells were cross-linked with 1% formaldehyde, which was added to the medium for 15 min at room temperature. After a quenching step with 0.125 M glycine, fixed cells were washed twice with PBS containing 1 μM PMSF and protease inhibitors, pelleted and lysed in lysis buffer (1% SDS, 10 mM EDTA and 50 mM Tris-HCl, pH 8.1) at 2 × 10<sup>7</sup> cells/ml. 10<sup>7</sup> cells (equivalent to 40–50 μg of chromatin) were used per immunoprecipitation reaction with 25 μg of antibody. Sonication was performed with a Covaris system (shearing time 30 min, 20% duty cycle, intensity 6, 200 cycles per burst, and 30 s per cycle) in a minimum volume of 2 ml. For calibrated ChIP-seq in siC-, siSA1- and siSA2-treated MCF10A cells, 20% of chromatin from mouse embryonic stem (ES) cells was added to the human chromatin. We doubled the amount of antibody used for the immunoprecipitation reactions to reduce differences on antibody saturation among conditions. ChIP-seq profiles for each antibody were multiplied by the occupancy ratio (OR) = (W<sub>m</sub>IP<sub>h</sub>)/(W<sub>i</sub>IP<sub>m</sub>), where W<sub>h</sub> and IP<sub>h</sub> are the number of reads mapped to the human genome from input (W) and immunoprecipitated (IP) fractions, and W<sub>m</sub> and IP<sub>m</sub> are reads mapped to the mouse genome from the input and IP fractions<sup>55</sup>.

Immunoprecipitated chromatin (6–10 ng, as quantified by fluorometry) was electrophoresed on an agarose gel, and independent sample-specific fractions of 100–200 bp were taken. An adaptor-ligated library was completed by limited-cycle PCR with Illumina PE primers (11–13 cycles). DNA libraries were applied to an

Illumina flow cell for cluster generation and sequenced on the Illumina Genome Analyzer Iix (GAIIx). Image analysis was performed with Illumina Real Time Analysis software (RTA1.8).

Alignment of 50-bp-long (76-bp-long for calibrated ChIP samples) sequences to the reference genome (GRCh37/hg19, February 2009) was performed using 'BWA and Bowtie2'<sup>56</sup> under default settings. Duplicates were removed using Picardtools (version 1.60), and peak calling was carried out using MACS2 (version 2.1.1.20160309) after setting the *q* value (FDR) to 0.05 or 0.01 (*SMC1*, *STAG1* and *STAG2* in HMECs) and using the '--extsize' argument with the values obtained in the 'macs2 predict' step<sup>57</sup>. All comparisons used the input tracks as 'control' and each one of the datasets as 'treatment'.

Common, SA1-only and SA2-only positions were defined using BEDtools v2.26, with a minimum of 1-nt overlap. Common positions were defined in two steps. (i) Overlap between SMC1 and SA1 bed files was performed by using the '-wa -wb' argument, and the positions obtained were concatenated and sorted by using the 'cat' and 'sort -k1,1 -k2,2n' commands. The output was merged by using the 'bedtools merge' function and was considered as one dataset. (ii) This was overlapped with the SA2 dataset as described above. SA1-only and SA2-only positions were those in which SA1 or SA2 did not overlap with each other.

Mean read-density profiles and read-density heat maps for different chromatin-binding proteins were generated with deepTools 2.0 (ref. <sup>58</sup>) BAM files of processed reads and plotting them around peak summits of SA1-only, SA2-only or common positions.

For motif discovery analysis, whole sequences of cohesin positions were extracted and used for motif enrichment analysis using MEME-ChIP from MEME<sup>59</sup>. Default parameters were used except for the following ones: -ccut 0, -meme-mod anr, -meme-minw: 6, -meme-maxw: 50, -nmeme: 600, -meme-nmotifs: 10, -meme-maxsize: 200,000.

Enrichment of cohesin positions (SA1-only, SA2-only and common) at HMEC and HCAEC chromatin states<sup>30</sup> was defined by using the 'intersect' function from BEDtools utilities (v2.26), with a minimum of 1-nt overlap. The analysis was performed making sure that one position did not belong to two different chromatin states.

To analyze cohesin distribution along super-enhancers, ChIP-seq reads from SA1 and SA2 in HMECs and HCAECs were plotted along HMEC super-enhancers<sup>32</sup> using the 'scale-regions' parameter from deepTools to adjust all of the super-enhancers to a predefined size and applying a local regression (LOESS) to smooth the read signals.

**Chromatin immunoprecipitation-qPCR and Re-ChIP.** ChIP-qPCR on immunoprecipitated chromatin was performed using the SYBR Green PCR Master Mix and an ABI Prism 7900HT instrument (Applied Biosystems). Primers were designed using OligoPerfect Designer (Invitrogen), and reactions were performed in triplicate. Chromosome coordinates of the validated peaks and the corresponding primers are listed in Supplementary Table 1. The relative amount of each amplified fragment was normalized with respect to the amplification obtained from input DNA using the ΔΔC<sub>t</sub> method and is represented as indicated in the corresponding figure legends.

The Re-ChIP experiment was performed with the Re-ChIP-IT kit (cat. no. 53016, Active Motif) according to the manufacturer's protocol. Briefly, MCF10A cells were fixed, lysed and sonicated as described in the ChIP protocol. 50 μg of chromatin was incubated with 20 μg of the first antibody (anti-SA1, anti-SA2 or IgG) in the presence of magnetic beads. The beads were washed, and the material was eluted and further incubated with 5 μg of the second antibody (anti-SA1, anti-SA2, anti-SMC1 or IgG). Eluted chromatin was analyzed by qPCR. 1 ng of the immunoprecipitated chromatin from two conditions—SA2-specific ChIP followed by IgG Re-ChIP and SA2-specific ChIP followed by anti-SA1 Re-ChIP—was used to prepare libraries for Re-ChIP sequencing. Libraries were prepared with 18 PCR cycles. Peaks were called in the SA2-SA1 Re-ChIP experiment after normalization with SA2-IgG Re-ChIP signals.

**qRT-PCR and RNA sequencing.** cDNAs were prepared with the Superscript II reverse transcriptase (Invitrogen) from total RNA (RNeasy Mini Kit, Qiagen), and qRT-PCR analyses were performed using the SYBR Green PCR Master Mix and an ABI Prism 7900HT instrument (Applied Biosystems). Primers (Supplementary Table 1) were designed using OligoPerfect Designer (Invitrogen). Reactions were performed in triplicate. Expression was normalized to that of the endogenous housekeeping gene *GAPDH*, using the ΔΔC<sub>t</sub> method.

For RNA-seq libraries (three replicates for the condition), poly(A)<sup>+</sup> RNA was purified with the Dynabeads mRNA purification kit (Invitrogen) from DNase I-treated total RNA, randomly fragmented, converted to cDNA and processed through subsequent enzymatic treatments of end repair, dA-tailing, and ligation to adaptors as per Illumina's protocol (TruSeq RNA Sample Preparation Guide; Part 15008136 Rev. A). The adaptor-ligated library was completed by limited-cycle PCR with Illumina PE primers (8 cycles). The resulting purified cDNA library was applied to an Illumina flow cell for cluster generation (TruSeq cluster generation kit v5), and it was sequenced on the Genome Analyzer Iix with SBS TruSeq v5 reagents by following the manufacturer's protocols. Fastq files with 50-nt single-end sequenced reads were quality-checked with FastQC (S. Andrews,

<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and aligned to the human genome (GRCh37/hg19) with Nextpresso (<http://bioinfo.cnio.es/nextpresso/>) executing TopHat-2.0.0 using Bowtie 0.12.7 and Samtools 0.1.16 allowing two mismatches and five multi-hits. Transcript assembly, estimation of their abundances and differential expression were calculated with Cufflinks 1.3.0 using the human genome annotation dataset GRCh37/hg19 from Ensembl. To account for multiple-hypothesis testing, the estimated significance level ( $P$  value) was adjusted using Benjamini–Hochberg FDR correction. For differential expression, FDR < 0.05,  $\log_2$  (fold change) < -0.5 or  $\log_2$  (fold change) > 0.5, and FPKM > 3, in at least one of the two conditions compared, was required.

GSEAPreranked was used to perform a gene set enrichment analysis<sup>60</sup>. We used the RNA-seq gene list ranked by statistic, setting ‘gene set’ as the permutation method, and we ran it with 1,000 permutations.

**Hi-C analysis.** MCF10A cells were arrested in the G1 stage of the cell cycle by means of high-confluency culture (150,000 cells/cm<sup>2</sup>). Hi-C was performed as described<sup>42</sup> using the MboI enzyme. Two library replicates per condition were sequenced (>200 million reads each; Supplementary Dataset 7). Data were processed using TADbit<sup>61</sup> for read quality control, read mapping, interaction detection, interaction filtering and matrix normalization. First, the reads were checked by using an implemented FastQC protocol in TADbit. This allowed discarding problematic samples and detection of systematic artifacts. Then, we used a fragment-based strategy in TADbit to map the remaining reads to the reference human genome (GRCh38). The mapping strategy resulted in ~80% of reads being mapped uniquely to the genome. Next, we filtered non-informative contacts between two reads—including self-circles, dangling-ends, errors, random breaks or duplicates. The final interaction matrices resulted in 272–303 million valid interactions per experimental condition (Supplementary Dataset 7). These valid interactions were then used to generate genome-wide interaction maps at 100 kb and 40 kb to segment the genome into the so-called A–B compartments and TADs, and to produce differential interaction maps.

A–B compartments were calculated by using vanilla-normalized and decay-corrected matrices as implemented in TADbit. Briefly, compartments were detected by calculating the first component of a principal-component analysis (PCA) of chromosome-wide matrices and assigning A compartments to the genomic bin with positive PCA1 values and high gene density (Fig. 6b). Conversely, B compartments were assigned to the genomic bin with negative PCA1 values and low gene density. TADs were identified by using 40-kb resolution vanilla-normalized and decay-corrected matrices as input to the TAD detection algorithm implemented in TADbit. TAD border localization, as well as strength, was calculated and used to identify conserved borders and their strength (Fig. 6c–e). A border was considered to be conserved between siControl and siSA1 or siSA2 experiments if it was localized within  $\pm 2$  bins in both experiments. Box plots were generated with the Python plotting library Matplotlib. Raw matrices normalized by coverage (i.e., all three experiments were scaled to have the same number of final valid interactions) at 100-kb resolution were also used for studying Hi-C interactions as a function of genomic distance. This genomic decay was obtained for each chromosome to a maximum genomic distance of 50 Mb, and the average was then calculated to obtain a genome-wide curve in siSA1 and siSA2 experiments (Fig. 6f). The same 100-kb matrices were used to determine differential Hi-C interactions between siControl and siSA1 or siSA2 experiments (Fig. 6g). These differential interactions were then classified according to their compartment localization and intra-TAD or inter-TAD distribution (Fig. 6h). Finally, the enrichment or depletion of genes (represented by their transcription start site), RNA (based on RNA-seq data), and CTCF- and cohesin-binding sites (SA1-only, SA2-only and common) was analyzed by a log odds analysis of observing such features in genomic bins belonging to A and B compartments, A–B borders or TAD borders (Fig. 6i). The log odds distributions were assessed for their distribution being statistically different than zero, as for a Fisher's exact test ( $P < 0.005$ ). The TADbit software used for Hi-C analyses is freely available as a Github repository at <https://github.com/3DGenomes/tadbit>.

**Immunoprecipitation and LC-MS/MS analysis.** Whole-cell extracts from MCF10A cells were prepared by lysis on ice for 30 min in TBS supplemented with 0.5% NP-40, 0.5 mM DTT, 0.1 mM PMSF and 1× complete protease inhibitor cocktail (Roche) followed by sonication. NaCl was added to 0.3 M, and the extract was rotated for 30 min at 4 °C. After centrifugation, the soluble fraction was recovered and diluted to bring the extract back to 0.1 M NaCl, and 10% glycerol

was added. Antibodies were cross-linked to protein A Pureproteome magnetic beads (Millipore) at 1 mg/ml (anti-SA1, anti-SA2 and IgG (as control)) and incubated with extracts overnight at 4 °C. The beads were washed six times with 20 volumes of lysis buffer, and proteins were eluted in two consecutive steps in two volumes of elution buffer (8 M urea, 100 mM Tris-HCl, pH 8) by shaking for 10 min. Samples were digested by standard filter-aided sample preparation (FASP)<sup>62</sup>. Proteins were reduced with 10 mM DTT, alkylated with 50 mM iodoacetic acid for 20 min in the dark and digested with 1:50 Lys-C (Wako) for 4 h. Samples were diluted in 50 mM ammonium bicarbonate and digested with 1:100 trypsin (Promega) overnight at 37 °C. The resulting peptides were desalted by using a Sep-Pak C18 cartridge for SPE (Waters Corp.), vacuum-dried and resuspended in 0.5% formic acid. Immunoprecipitates were analyzed with a nanoLC Ultra system (Eksigent) coupled with a LTQ-Orbitrap Velos instrument (Thermo) via nanoESI (ProxeonBiosystem). Two technical replicates were performed. Raw data were analyzed using MaxQuant 1.5.3.30<sup>63</sup> with Andromeda<sup>64</sup> as the search engine against the UniProtKB and Swiss-Prot databases (20,584 sequences). Peptides were filtered at 1% FDR. For protein assessment (FDR < 1%), at least one unique peptide was required for both identification and quantification. Other parameters were set as default. The resulting ‘protein\_group.txt’ file was loaded in Perseus<sup>65</sup> (v1.5.1.6). Missing values were imputed from a normal distribution. A two-sample Student's  $t$  test (one sided) was used, corrected for multiple testing by using a permutation-based approach.

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** ChIP-seq, RNA-seq and Hi-C data from this study have been submitted to the GEO database (GSE101921). Additional source data are available upon reasonable request.

## References

- Carretero, M., Ruiz-Torres, M., Rodríguez-Corsino, M., Barthelemy, I. & Losada, A. PDS5B is required for cohesion establishment and aurora B accumulation at centromeres. *EMBO J.* **32**, 2938–2949 (2013).
- Gocke, C. B. & Yu, H. ZNF198 stabilizes the LSD1–CoREST–HDAC1 complex on chromatin through its MYM-type zinc fingers. *PLoS One* **3**, e3255 (2008).
- Méndez, J. & Stillman, B. Chromatin association of human origin recognition complex, CDC6 and minichromosome maintenance proteins during the cell cycle: assembly of pre-replication complexes in late mitosis. *Mol. Cell. Biol.* **20**, 8602–8612 (2000).
- Hu, B. et al. Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res.* **43**, e132 (2015).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
- Zhang, Y. et al. Model-based analysis of ChIP-seq (MACS). *Genome Biol.* **9**, R137 (2008).
- Ramírez, F. et al. deepTools2: a next-generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**(W1), W160–W165 (2016).
- Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
- Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* **23**, 3251–3253 (2007).
- Serra, F. et al. Automatic analysis and 3D-modeling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **13**, e1005665 (2017).
- Wiśniewski, J. R., Zougman, A., Nagaraj, N. & Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362 (2009).
- Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
- Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
- Tyanova, S. et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740 (2016).

## 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 mycoplasma-free. MCF10A cells were periodically tested for mycoplasma (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.*

## 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: ybatoyqivfmzsf

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](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 replicates 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

6. Describe the sequencing depth for each experiment.

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).

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

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

\*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.

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 tumorigenesis 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 predicted step)' and using input as the control.