

Published in final edited form as:

Nature. 2020 February ; 578(7795): 472–476. doi:10.1038/s41586-019-1910-z.

The structural basis for cohesin-CTCF anchored loops

Yan Li^{#1}, Judith H.I. Haarhuis^{#2}, Ángela Sedeño Cacciatore^{#2}, Roel Oldenkamp², Marjon S. van Ruiten², Laureen Willems², Hans Teunissen^{2,3}, Kyle W. Muir^{1,4,*}, Elzo de Wit^{2,3,*}, Benjamin D. Rowland^{2,*}, Daniel Panne^{1,5,*}

¹European Molecular Biology Laboratory, 38042 Grenoble, France ²Division of Gene Regulation, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

³Oncode Institute ⁴MRC Laboratory of Molecular Biology, Cambridge, United Kingdom ⁵Leicester Institute of Structural and Chemical Biology, Department of Molecular and Cell Biology, University of Leicester, Lancaster Road, Leicester, LE1 7RH, United Kingdom

These authors contributed equally to this work.

Summary

Cohesin catalyzes folding of the genome into loops that are anchored by CTCF¹. The molecular mechanism of how cohesin and CTCF structure the 3D genome has remained unclear. Here we show that a segment within the CTCF N-terminus interacts with the SA2-SCC1 subunits of cohesin. A 2.6Å crystal structure of SA2-SCC1 in complex with CTCF reveals the molecular basis of the interaction. We demonstrate that this interaction is specifically required for CTCF-anchored loops and contributes to the positioning of cohesin at CTCF binding sites. A similar motif is present in a number of established and novel cohesin ligands, including the cohesin release factor WAPL^{2,3}. Our data suggest that CTCF enables chromatin loop formation by protecting cohesin against loop release. These results provide fundamental insights into the molecular mechanism that enables dynamic regulation of chromatin folding by cohesin and CTCF.

Keywords

Cohesin; CTCF; TAD; Chromatin loops; 3D genome; SA2

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

*Correspondence and requests for materials should be addressed to D.P. (daniel.panne@le.ac.uk) or B.D.R. (b.rowland@nki.nl). Or E.d.W. (e.d.wit@nki.nl) or K.W.M. (kmuir@mrc-lmb.cam.ac.uk).

Data availability Coordinates are available from the Protein Data Bank under accession number 6QNX for the SA2-SCC1-CTCF complex. The generated Hi-C, RNA-Seq and ChIP-Seq data has been deposited in GEO (accession number GSE126637).

Author Contributions Y.L. and K.W.M. initiated the project and proposed the CES motif. Y.L. performed biochemical studies and structural analyses with support from K.W.M., J.H.I.H., R.O., M.S.V.R., L.W. and H.T. performed wet-lab cell-based experiments and A.S.C. performed bioinformatic analyses. K.W.M., E.d.W., B.D.R. and D.P. provided supervision. Y.L., K.W.M., B.D.R. and D.P. were involved in conceptualization, project administration and wrote the original and revised draft with input from all authors.

Competing Interests The authors declare no competing interests

Author Information Reprints and permissions information is available at www.nature.com/reprints. Readers are welcome to comment on the online version of the paper.

Introduction

The interphase genome is folded in 3D through the concerted action of cohesin and CTCF. These architectural factors regulate the interactions between regulatory elements along chromosomes to control gene expression^{1,4,5}. Cohesin is thought to catalyze genome folding through a process known as ‘loop extrusion’, which involves the formation of chromosome loops that are progressively enlarged^{6–10}. Genomic regions within which cohesin forms loops are also known as topologically associating domains (TADs), or loop domains. TADs are flanked by CTCF sites that are thought to act as barriers to the loop extrusion process^{11,12}. CTCF only acts as such a boundary when the 3’ ends of CTCF binding motifs are oriented towards the inside of the TAD^{9,13,14}; Consequently, only convergently-oriented pairs of CTCF sites form CTCF-anchored loops^{15,16}.

This model is supported by genetic manipulation of cohesin and CTCF. Depletion of the core cohesin subunit SCC1 leads to loss of TADs^{12,17}. In contrast, depletion of the cohesin release factor WAPL increases the size of chromatin loops^{10,12,18}. CTCF depletion leads to a dramatic loss of CTCF-anchored loops^{11,12}. How CTCF can act as a directional boundary that controls cohesin loop extrusion, however, remains unknown.

Here we have investigated the mechanism of cohesin interaction with CTCF, and how this interaction contributes to genome organization. We have identified an N-terminal segment of CTCF that directly engages the SA2-SCC1 subcomplex of cohesin. A crystal structure of the SA2-SCC1-CTCF complex elucidates the molecular basis of the interaction. Mutation of key amino acids in the interface abolishes CTCF-anchored loops, but only partially impairs accumulation of cohesin at CTCF binding sites across the genome. Thus, in addition to its function as a translocation barrier, CTCF possesses a distinct loop stabilizing activity, which is realized through a direct interaction with cohesin. Furthermore, we observe intermolecular competition between CTCF and the cohesin release factor WAPL for this interface, suggesting a mechanism by which chromatin loop formation may be dynamically regulated.

Results

Structure of the SA2-SCC1-CTCF complex

Previous data indicate that CTCF directly interacts with the SA2 subunit of the cohesin complex^{19,20}. To map this interaction, we produced a series of CTCF truncations as GST fusion proteins and performed pulldown assays against a complex of SA2 and SCC1². CTCF fragments containing amino acids 227-235 generally retained SA2-SCC1 on GST-beads (Extended Data Fig. 1a,b). Isothermal calorimetry experiments further showed that the interaction is largely driven by amino acids 222-231 of CTCF as it retained an equilibrium dissociation constant ($K_d = 1.04 \pm 0.20 \mu\text{M}$) comparable to an extended CTCF construct ($K_d = 0.62 \pm 0.07 \mu\text{M}$; Extended Data Fig. 1c and Extended Data Table 1a). To understand the molecular details, we produced crystals of the SA2-SCC1 complex in the presence of a peptide comprising the CTCF binding motif and determined the structure by molecular replacement at a resolution of 2.6 Å (Extended Data Table 1b). An Fo-Fc omit electron density Fourier map exhibited clear features corresponding to the CTCF peptide (Extended Data Fig. 1d).

The CTCF peptide is bound to the convex surface of SA2 (Fig. 1a,b). The CTCF binding surface is predominantly hydrophobic and composed of amino acids contributed by both SA2 and SCC1. The lead ‘anchoring’ amino acids of CTCF, which bury the largest solvent-accessible surface area upon binding, are Y226 and F228 (Fig. 1b). F228 inserts into a pocket comprising amino acids from SCC1 (S334, I337, L341) and SA2 (Y297, W334) (Extended Data Fig. 1f). The hydroxyl group of Y226 hydrogen bonds with D326 of SA2 in a deep hydrophobic pocket lined by L329, L366 and F367 (Fig. 1c,d). E229 and E230 of CTCF constitute secondary anchoring residues, which presumably contribute to binding specificity by forming salt bridges with R298 of SA2 and R338 of SCC1. As CTCF engages a composite binding surface containing amino acids from SCC1 and SA2, prior mapping studies using isolated SA2 may have been misleading²⁰.

Analysis of the CTCF binding interface

Mutagenesis of Y226A or F228A in CTCF abolished SA2-SCC1 binding in a GST pulldown assay (Fig. 1e). Likewise, mutation of critical amino acid residues in SA2, including W334A, F371A and F367A or in SCC1 I337A/L341A, abolished CTCF binding (Fig. 1f). SA2 contains an 86 amino acid motif, termed the ‘stromalin conservative domain’^{21,22} or ‘conserved essential surface’ (CES)^{2,23}, which is conserved from fungi to mammals and coincides with the CTCF binding pocket. For simplicity, we will henceforth refer to the composite SA2-SCC1 binding pocket as the ‘CES’. Mapping of sequence conservation onto the structure confirms that the CES is highly conserved (Fig. 1g, Extended Data Fig. 2a). A series of missense mutations are found in SA2, SCC1 and CTCF in various cancer types²⁴. Mapping of mutation frequencies onto the structure shows that amino acids that are largely buried in the interface are hotspots in cancer (Extended Data Fig. 2b).

Previous data indicate that the SA2-SCC1 complex interacts with multiple cohesin regulators^{2,23,25}. This includes two factors with opposing functions: WAPL, the general cohesin release factor, and Shugoshin (SGO1), a factor crucial for the protection of centromeric cohesion during mitosis^{2,26–28}. This antagonism arises as a result of direct competition for binding to the CES of SA2-SCC1². As mutants reported to interfere with both SGO1 and WAPL binding cluster in the CES, we investigated whether these proteins bind to SA2-SCC1 by a mechanism comparable to that of CTCF. In SGO1, the reported CES-binding domain (amino acids 313- 353) contains a conserved FGF motif which strongly resembles that of the CTCF peptide. Vertebrate WAPL also contains several FGF motifs in its N-terminal region which are potentially involved in cohesin regulation^{3,29}. A minimal fragment of WAPL capable of competing with SGO1 for access to the CES (amino acids 410-590) contains two such FGF motifs². We observed that a peptide spanning the second and third FGF motif of WAPL (amino acids 423-463) bound to SA2-SCC1 with a K_d of ~32.8 μ M (Extended Data Fig. 2c) whereas a peptide comprising only the third motif bound more weakly (Extended Data Table 1a). The peptide containing the CTCF CES motif therefore binds with higher affinity than do peptides containing the WAPL motif(s).

CTCF stabilizes cohesin on chromatin

The observation that CTCF and WAPL can bind to the same surface on SA2-SCC1 raises the possibility that their interaction with the CES is mutually exclusive (Fig. 2a). To

determine whether WAPL competes with CTCF for binding to the SA2-SCC1 CES, we performed GST-pulldown competition assays. Indeed, titration of WAPL residues 1-600 against a preformed complex of GST-CTCF and SA2-SCC1 depleted the latter from the beads (Fig. 2b). Similarly, titration of a SGO1 phospho-T346 peptide, previously reported to preclude WAPL binding², also displaced SA2-SCC1 from GST-CTCF (Extended Data Fig. 2d). Hence, the CES of SA2-SCC1 is a general interaction hub for multiple cohesin regulators (Extended Data Fig. 2e). Whereas SGO1 precludes WAPL binding, thus stabilizing centromeric cohesin in mitosis, CTCF could exert a similar function at CTCF sites in interphase.

To test whether the CTCF-CES interaction stabilizes cohesin on chromatin, we mutated the endogenous allele of CTCF in the human haploid HAP1 cell line using CRISPR/Cas9 technology. We thereby obtained HAP1 cells that harboured Y226A F228A mutant CTCF as their sole copy of CTCF (Extended Data Fig. 3a,b). These cells displayed no obvious proliferation defects. To study the consequences of the CTCF mutations on cohesin turnover on chromatin, we endogenously tagged the core cohesin subunit SCC1 with a HaloTag in both wild type and CTCF Y226A F228A cells (Extended Data Fig. 3c), and performed fluorescence recovery after photobleaching (FRAP) experiments. In wild-type cells, we found that, over a period of 20 minutes a fraction of the fluorescent cohesin population did not recover. In CTCF Y226A F228A cells, however, we observed a near-complete recovery by FRAP, demonstrating that cohesin is more mobile in these cells (Fig. 2c,d). The CTCF-CES interaction therefore stabilizes a sub-population of cohesin on chromatin.

CTCF-CES interaction is required for CTCF-anchored loops

To investigate the role of the cohesin-CTCF interaction in chromosome organization, we generated Hi-C profiles of wild type and CTCF Y226A F228A mutant cells. Wild-type HAP1 cells displayed clear loops connecting CTCF sites (Fig. 3a, and Extended Data Fig. 4), however Hi-C matrices of CTCF Y226A F228A mutant cells revealed a robust ablation of CTCF-anchored loops (Fig. 3a). By systematically scoring the number of loops we found that the CTCF Y226A F228A mutation indeed led to a loss of the vast majority of detectable loops across the genome (from 2756 in the wild-type to 98 in the mutant cells) (Fig. 3b). An aggregate peak analysis, which quantifies the contact frequency of all the loops identified in wild-type cells, likewise showed a dramatic loss of these contacts (Fig. 3c and Extended Data Fig. 4d).

CTCF sites not only lie at the bases of CTCF-anchored loops, but also form the boundaries of TADs (Extended Data Fig. 4a). We then assessed the effect of the CTCF Y226A F228A mutation on TADs, and found that these structures were, to a considerable degree, still present in CTCF Y226A F228A cells, but have less-clear edges (Fig. 3a). Aggregate TAD analysis further confirmed that TAD-like structures do exist in CTCF Y226A F228A cells, but indeed have less-clear boundaries (Extended Data Fig. 4b,c,e,f), and completely lack CTCF loops at their corners (Extended Data Fig. 4b,e). Our results therefore support the notion that in CTCF Y226A F228A cells cohesin can still form the loops along DNA that make up the contacts within TADs, but that cohesin is not stabilized at CTCF sites to allow for the formation or maintenance of CTCF-anchored loops.

CTCF interaction promotes cohesin localization to CTCF sites

To assess whether the CTCF-CES interaction affects cohesin abundance at loop anchors, we performed quantitative chromatin immunoprecipitation experiments (ChIP-qPCR). We selected CTCF sites at the base of loops (Fig. 4a and Extended Data Fig. 5a) and found that the CTCF Y226A F228A mutation reduced the abundance of cohesin at the majority of these loci. In contrast, cohesin levels at a nearby locus that did not contain a CTCF site were not affected (Fig. 4b, and Extended Data Fig. 5b). CTCF binding to the corresponding CTCF sites was also largely unaffected (Extended Data Fig. 5c-e). We then assessed cohesin distribution genome-wide by ChIP-Seq and found that the Y226A F228A mutation decreased cohesin localization to CTCF sites, but had little to no effect on cohesin localization at unrelated sites (Fig. 4c,d). While the CTCF Y226A F228A mutations reduced cohesin levels at CTCF sites, cohesin was to a significant degree still present at CTCF sites. Our data therefore support the model that CTCF influences cohesin in two distinct ways: i) it halts cohesin at CTCF sites and ii) it stabilizes cohesin at the base of CTCF-anchored loops. The former function could be important for defining TAD boundaries, while the binding of CTCF to the cohesin CES affects this latter function, and may thereby prevent the disruption of CTCF-anchored loops.

To evaluate the consequences of the loss of CTCF-anchored loops on gene expression, we performed RNA-Seq analyses. The CTCF Y226A F228A mutation affected the expression of >2000 genes. While comparable numbers of genes were upregulated as were downregulated, the most strongly impacted genes were more frequently downregulated (Extended Data Fig. 7a). Thus, the Y226 F228 interface of CTCF, and by extension cohesin-CTCF anchored loops, are apparently key to proper expression of these genes. Despite this effect on gene expression and the loss of virtually all CTCF-anchored loops, cells harbouring only this mutant form of CTCF are viable. Previously, CTCF has been shown to be essential for viability of murine embryonic stem cells¹¹. We therefore tested whether CTCF is essential for the viability of HAP1 cells, and found that siRNA-mediated CTCF depletion was lethal to both control HAP1 cells and CTCF Y226A F228A mutant cells (Extended Data Fig. 7b,c). Thus, CTCF has essential roles that are apparently independent of CES engagement and the formation of CTCF-anchored loops in these cells.

Identification of CES ligands

To investigate the prevalence of the CES-binding factors, we compiled an alignment of known cohesin partners and derived a regular expression motif (Extended Data Fig. 2e,f). We used this motif to query the human and budding yeast proteomes for proteins containing similar binding motifs³⁰. From the set of nuclear proteins arising from this search, we were able to identify known cohesin regulators as well as several novel potential binding factors. We generated peptide arrays bearing these sequences and assayed binding of SA2-SCC1, employing an SA2 (F371A)-SCC1 mutant complex as a negative control. We observed clear signal for the CTCF peptide spanning 222-231, which was abolished in the SA2 (F371A)-SCC1 mutant (Extended Data Fig. 7d,e). A CTCF Y226F mutant showed ~1.5 fold reduced binding, apparently due to loss of the hydrogen bond between the hydroxyl group of CTCF Y226 and D326 of SA2 (Extended Data Table 2). Consistent with our pull-downs, the CTCF Y226A, F228A, and Y226A/F228A peptide variants failed to retain SA2-SCC1. The WAPL

peptides showed considerably weaker binding as compared to CTCF, while we could not detect binding for ligands such as SGO1 (see Methods and Extended Data Table 1a). Robust binding was observed for MCM3, a subunit of the replicative helicase, SYCP3, a component of the synaptonemal complex, ZGPAT, a transcriptional repressor, and CENPU, a subunit of the inner kinetochore. Thus, SA2-SCC1 CES potentially facilitates cohesin regulation for a number of functionally divergent chromosomal processes.

Discussion

Our study reveals that CTCF binds to a conserved essential surface (CES) on the SA2-SCC1 cohesin subcomplex. The ablation of this interaction results in a near-complete loss of CTCF-cohesin anchored loops. Thus, CTCF does not simply present a passive barrier to cohesin-mediated loop extrusion, but specifically interacts with the CES to stabilize cohesin at these loci, and to prevent loop disruption. Accordingly, impairment of the CTCF-CES interaction renders cohesin more dynamic (Fig. 2c,d).

SA2, SCC1 as well as CTCF are frequently mutated in a number of different tumour types³⁹ and the mutations cluster in the CES (Extended Data Fig. 2b). Therefore, the dysregulation of chromatin looping may be causally related to carcinogenesis (Extended Data Fig. 7a)^{31,32}.

We envisage two alternative scenarios for the formation of CTCF-anchored chromatin loops. In the first model (Fig. 4e), cohesin initiates loop enlargement at distal chromatin loci. These cohesin complexes remain dynamic because the cohesin release factor WAPL directly binds to cohesin by engaging the CES^{2,3,29} and PDS5^{12,27,33}, and promotes opening of cohesin rings at the SMC3-SCC1 interface³⁴⁻³⁶. Alternatively, loop enlargement commences at CTCF sites³⁷. Cohesin then catalyzes DNA looping at these sites because CTCF counteracts DNA release (Fig. 4f). These models are not necessarily mutually exclusive, as a cohesin complex that initiates looping in the former mode may well be converted into the latter upon encountering CTCF. As CTCF directly competes with WAPL for binding to the CES (Fig. 2a,b), we propose that this interaction stabilizes chromatin loops.

We propose a model for how cohesin and CTCF co-associate on DNA (Fig. 4g). Our model indicates that cohesin only engages CTCF when approaching its N-terminus. Specifically, the 34 amino acid flexible linker connecting the YxF motif to the first DNA-binding zinc finger of human CTCF, is sufficiently long to allow SA2-SCC1 DNA binding towards the N-, but not the C-terminus of CTCF (Fig. 4g), thus confirming previous mapping studies³⁸. Stabilization of cohesin by engagement of the CTCF N-terminus may in fact explain why TAD boundaries arise preferentially when CTCF binding sites are convergently oriented^{9,13-16,38,39}. If an individual cohesin complex anchors itself at the N-terminus of CTCF, and then reels in DNA until it encounters a cohesin that is likewise reeling from the opposite CTCF site, this would bring together CTCF sites³⁷. Loop formation by the related condensin complex appears to involve a DNA anchoring function of its HEAT-repeat subunit Ycg1, a paralogue of SA2/ScC3⁴⁰⁻⁴². These different complexes may therefore use a similar anchoring principle to build loops and provide structure to genomes. As the CES interface is conserved between SA isoforms, we anticipate that ligand binding will affect all cohesin

variants in a similar manner. Similarly, this interface is also conserved through Scc3 in lower Eukaryotes, despite the absence of CTCF in such organisms. The CES therefore is likely to represent an ancient interaction hub on cohesin.

The observation that CTCF-CES interaction controls DNA looping indicates that this aspect of cohesin function can be regulated by F/YxF motif containing cohesin ligands. A number of other genome regulatory factors contain F/YxF motifs including Shugoshin (Extended Data Fig. 7d,e), which protects centromeric chromatid cohesion by antagonizing WAPL binding to the CES of SA2-SCC1². We therefore predict that a number of proteins containing F/YxF motifs engage the CES and thereby modulate cohesin's ability to catalyse genome folding in functionally divergent chromosomal processes.

Methods

Constructs, protein expression and purification

The human SA2 fragment 80-1060, cloned into pGEX-6P and codon-optimized for expression in *Escherichia coli*, was obtained from Hongtao Yu (UT Southwestern, USA). The construct encodes an N-terminal GST tag and C-terminal SA2 separated by a PreScission protease cleavage site. A plasmid encoding SCC1 was obtained from Jan-Michael Peters (IMP, Vienna). SA2 was co-expressed with an N-terminally 6xHis-tagged fragment of SCC1 spanning residues 281-420 cloned into the NcoI-NotI sites of a pACYCDuet-1 vector (Merck Millipore). CTCF constructs were cloned into the BamHI and NotI sites of pGEX-6P1. Mutagenesis was performed using a QuikChange Lightning site-directed mutagenesis kit (Agilent). All the proteins were expressed in *E. coli* BL21(DE3) by autoinduction⁴³. Cells were grown at 37°C until OD_{600nm} = 0.6 and then shifted to 18°C for 16 hours. Cells were harvested with a JLA-8.1 rotor (Beckman) and washed once with ice-cold PBS buffer. Pellets were resuspended in buffer 1 (40 mM TRIS, pH 7.5, 500 mM NaCl, 0.5 mM TCEP), lysed using a microfluidiser (Microfluidics) and centrifuged at 4°C for 1h at 15000 rpm using JA-20/14 rotors (Beckman).

The GST- and His-tagged SA2-SCC1 complex was applied to Co²⁺ conjugated IMAC Sepharose resin (GE Healthcare) using a Minipuls3 peristaltic pump (Gilson), washed with buffer 1 supplemented with 20 mM Imidazole and eluted using buffer 1 supplemented with 300 mM Imidazole. Co²⁺ eluate was then bound to Glutathione Sepharose 4 Fast Flow resin (GE Healthcare) using a Minipuls3 peristaltic pump (Gilson), washed with buffer 1, and eluted by adding 10 mM reduced L-Glutathione (Sigma-Aldrich) into buffer 1. The GST tag was cleaved by PreScission protease (EMBL core facilities) during overnight incubation at 4°C. Cleaved protein was concentrated using an Amicon Ultra -15 concentrator (Millipore) and applied to a MonoQ 5/50 GL column (GE Healthcare) in buffer 2 (40 mM TRIS, pH7.5, 150 mM NaCl, 0.5mM TCEP) and eluted via a linear gradient of buffer 2 containing 1M NaCl and further purification using a HiLoad 16/60 Superdex 200 prep-grade column (GE Healthcare) in buffer 3 (20 mM TRIS, pH7.7, 300 mM NaCl, 5 mM TCEP). The final purified proteins were concentrated using an Amicon Ultra -15 concentrator (Millipore) and flash frozen in liquid N₂ for storage at -80°C.

Crystallization and structure determination

Crystals of SA2 80-1060 in complex with SCC1 281-420 (otherwise denoted the SA2-SCC1 complex) were grown by hanging-drop vapor diffusion at 20°C by mixing equal volumes of protein at 8 mg ml⁻¹ and crystallization solution containing 0.06M Morpheus Divalents mix, 0.1M Morpheus buffer system 2, 48% (v/v) Morpheus EOD_P8K (Molecular Dimensions). Crystals were soaked for 24-48h with a peptide (peptid.de) including amino acid residues 222-231 of CTCF (Uniprot ID: Q8NI51; DVSVDYDFEEE). Crystals were cryo-protected by adding 15% glycerol to the well solution and flash frozen in liquid nitrogen.

Diffraction data for all crystals were collected at 100 K at an X-ray wavelength of 0.966 Å at beamline ID30A-1/MASSIF-1⁴⁴ of the European Synchrotron Radiation Facility, with a Pilatus3 2M detector using automatic protocols for the location and optimal centering of crystals⁴⁵. The beam diameter was selected automatically to match the crystal volume of highest homogeneous quality⁴⁶. Data were processed with XDS⁴⁷ and imported into CCP4 format using AIMLESS⁴⁸.

The structure was determined by molecular replacement using Phaser⁴⁹. A final model was produced by iterative rounds of manual model building in Coot⁵⁰ and refinement using PHENIX⁵¹. The CTCF-containing model was refined to 2.6 Å resolution with an R_{work} and an R_{free} of 25% and 27%, respectively (Extended Data Table 1b). Analysis by MolProbity⁵² showed that there are no residues in disallowed regions of the Ramachandran plot and the all atom clash score was 7.2. The model shown in Fig. 4f was generated by superposition on DNA of SA2-SCC1-CTCF (6QNX) with DNA-bound ySCC3-SCC1 (6H8Q)⁴² and a composite model of DNA-bound CTCF Zinc fingers assembled from 5YEF and 5YEL⁵³.

GST pulldowns and peptide arrays

For GST pulldowns, 10 μM GST-tagged CTCF constructs and 2.5 μM SA2-SCC1 were mixed in 50 μl buffer 4 (20 mM TRIS, pH7.7, 300 mM NaCl, 0.5 mM TCEP) + 0.1% Tween-20 containing 25 μl of a 50% slurry of GST Sepharose beads per reaction. For WAPL and SGO1 competition assays, 2.5 μM GST-tagged CTCF (86-267) was incubated with 1 μM of SA2-SCC1 and increasing concentrations of WAPL (1-600) or a T346-phosphorylated SGO1 peptide spanning 331-349 (molar ratios are indicated in each figure), under reaction conditions otherwise identical to GST pulldowns. Reactions were incubated at for 1h at 4°C. 25 μl of the reaction were withdrawn as the reaction input ('I') while the remainder was washed five times with 500 μl of buffer 4 + 0.1% Tween-20. Samples were boiled in 1 x SDS sample loading buffer (NEB) for 5 minutes to obtain the bound ('B') fraction, followed by SDS-PAGE analysis.

ITC was performed using a MicroCal iTC 200 (Malvern Panalytical) at 25°C. SA2-SCC1 and the CTCF, SGO1 and WAPL peptide ligands were dialyzed overnight at 4°C against 20 mM TRIS, pH7.7, 150 mM NaCl, 0.5 mM TCEP. For each titration, 300 μl of 50 μM SA2-SCC1 was added to the calorimeter cell. The concentration of peptides was adjusted to 500 μM and injected into the sample cell as 16 x 2.5 μl syringe fractions. Results were analysed and displayed using the Origin 7.0 software package supplied with the instrument. Data were analyzed using the one site binding model.

Peptide arrays, with an area of 3cm², were obtained from Rudolf Volkmer (immunologie.charite.de). Arrays were washed with 100% ethanol for 5 minutes on a shaker at 21°C, followed by 3 washes, for a total of 10 minutes in TBS-T buffer (50 mM Tris pH7.5, 150 mM NaCl, 0.05% Tween-20). For the blocking step, arrays were incubated in 1x blocking buffer (Sigma B6429) for 3 h at 21°C, followed by 3 washes in TBS-T for a total of 10 minutes. SA2-SCC1 and SA2 (F371A)-SCC1 were added to 1x blocking buffer at a final concentration of 1.2 µM and incubated with the array overnight at 4°C under gentle agitation. The membrane was washed 3 times (1x 30s, then 2x 5 minutes) at 21°C. The anti 6x anti-poly His-HRP antibody (Sigma A-7058) was diluted 1:2000 in 1x blocking buffer and incubated with the arrays for 1h at 21°C. The array was washed 3 times (1x 30s, then 2x 5 minutes) and developed by addition of 3,3'-Diaminobenzidine (Sigma D4293) for 1 minute followed by quenching in deionized H₂O. To measure non-specific binding of the anti 6xHistidine antibody, all steps were identical expect that no SA2-SCC1 protein solution was added to 1x blocking buffer during the overnight incubation step. Arrays were imaged with a BioRad Gel Doc XR+ Documentation system. Spot intensities were measured using ImageJ 1.52k. Three independent experiments were done and the apparent dissociation constants determined by normalization with ITC data from CTCF 222-231 (Extended Data Table 1a).

Genome editing and cell culture

Cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 10% FCS (Clontech), 1% Penicillin/Streptomycin (Invitrogen) and 1% UltraGlutamin (Lonza). The gRNA targeting exon 1 of CTCF was designed and annealed into pX330 (primer: 5'-CGATTTTGAGGAAGAAGCAGC-3'). To modify the targeted locus, we co-transfected a 120 basepair repair oligo containing the desired mutation and a silent mutation (repair oligo: 5' – CCAAAAAGAGCAAACACTGCGTTATACAGAGGAGGGCAAAGATGTAGATGTGTCTGT CGCCGATGCTGAAGAAGAAGCAGCAGGAGGGTCTGCTATCAGAGGTTAATGCAGAGAAAGTGGTTG-3'). pBabePuro was co-transfected in a 10:1 ratio to the pX330. Transfected clones were selected using 2 µg/ml puromycin for 2 days. Colonies were picked when they were clearly visible, gDNA of clones was isolated, and mutations were validated by Sanger sequencing.

To target the C-terminus of SCC1, a gRNA (primer: 5'-CCAAGTTCCATATTATATA-3') was cloned into px459 V2.0 (Addgene plasmid #62988). The SCC1-HaloTag HR template was a gift from James Rhodes⁵⁴. SCC1-Halo cell lines were generated by cotransfection of pX459 and the SCC1-HaloTag HR vector using FuGENE® HD Transfection Reagent. Cells were selected with puromycin (2 µg/ml) for 2 days. Colonies were picked when they were clearly visible and validated using western blot analysis and immunofluorescence.

Antibodies

The following antibodies were used for western blots: SMC1 (A300-055A, Bethyl), CTCF (07-729, Millipore) and ab128873, Abcam), HSP90 (F-8, Santa Cruz), SCC1 (05-908, Millipore) (Tubulin (T5168, Sigma), H4 (05-858, Millipore). All primary antibodies were used at a 1:1000 dilution, with the exception of HSP90 and Tubulin (1:10.000). Secondary

antibodies for western blot analysis were used in a 1:2000 dilution: Goat anti-Rabbit-PO and Goat anti-Mouse-PO (DAKO). For ChIPseq we used the following antibodies: SCC1 (ab992, Abcam), CTCF (3418S, Cell Signaling) and IgG (I5006, Sigma-Aldrich).

FRAP

Cells were grown on LabTekII-chambered cover glass (Thermo Scientific Nunc). Two days before imaging, cells were transfected with DHB-iRFP using FuGENE® HD Transfection Reagent. Prior to imaging, cells were incubated with 300 nM fluorescent HaloTag ligand JF585 for 30 min. Cells were washed three times with normal medium and incubated for 1 hour to allow exit of excess of ligand. Medium was replaced twice more with pre-warmed Leibovitz L-15 medium (Invitrogen). Live-cell imaging was performed on a Leica SP5 confocal microscope with a 63x 1.2 NA water objective using the LAS-AF FRAP-Wizard. Prior to bleaching 5 images were taken. Half of the nucleus of G1 cells was photobleached using 6 pulses of 100% transmission of a 561 laser. Subsequently, 600 frames were taken every 2 seconds. Fluorescence intensity was measured in the bleached and unbleached area by user-defined regions using ImageJ v1.52q and adjusted by hand for nucleus movement. Measurements were corrected for photobleaching by monitoring a non-bleached cell. Recovery was quantified by calculating the difference in intensity in the bleached and unbleached regions after background correction. Non-diffusive SCC1-Halo (Extended Data Fig. 3f), was quantified by the relative loss in fluorescence intensity in the unbleached region between the first frame post- and 5 frames pre-bleaching.

Colony formation assay

Cells were seeded at equal density and transfected with siRNA's targeting either no oligo, Luciferase, CTCF or SMC1. All siRNA's were ON-TARGETplus SMARTpools manufactured by Dharmacon. Transfection was repeated after 3 days and after an additional 4 days samples were fixed for 10 minutes with 96% Methanol and stained with 0.25% Crystal Violet. Cells treated by the same protocol were taken along for western blot analysis, samples were harvested 2 days prior to colony formation assay to have enough cells for western blot analysis.

Chromatin fractionation

For the chromatin fractionation experiment of Extended Data Fig. 3e, 50 million cells per cell line were harvested and fractionation was performed using Subcellular Protein Fractionation Kit for Cultured Cells (78840, ThermoFisher Scientific) according to manufacturer's protocol, with minor changes. The pellet was washed twice after centrifugation. Western Blots were performed as described¹⁰.

Hi-C

Samples for Hi-C were prepared as described previously¹⁰. Raw sequence data was mapped and processed using HiC-Pro v2.9⁵⁵ with hg19 as reference. Statistics on the number of valid pairs and percentage of cis contacts are summarized in Extended Data Table 3b,c. Replicates 1 and 2 are highly similar with a reproducibility > 0.98 as assessed by HiCRep v1.8.0⁵⁶ and were subsequently combined into one Hi-C dataset. The valid pair files

generated by HiC-Pro were used to create juicebox ready files using juicebox-pre (juicer tools v0.7.5)⁵⁷. For visualization, contact matrices were ICE normalized⁵⁸ and counts were normalized for 100 million contacts per sample.

Loops were then called with HICCUPS v1.11.09¹⁵ at 5kb, 10kb and 25kb resolutions. To visualize the genome-wide effect of the introduced CTCF mutations in loops we performed aggregate peak analysis¹⁵ as implemented in GENOVA v0.9.8 (<https://github.com/robinweide/GENOVA>), using previously defined loops in wild-type HAP1 cells¹⁰. Briefly, for a set of loop coordinates a square submatrix is selected such that it is centred on the corresponding coordinates, with a 100kb flanking region upstream and downstream. These submatrices are then averaged to obtain a mean contact map for these locations.

Similar to the aggregated peak analysis, aggregate TAD analysis was done to visualize how TAD structures are affected by the CTCF mutations. For this analysis we employed TADs previously defined for wild type HAP1 cells¹⁰. In short, these TADs were called using HiCseg on 10 kb matrices as input, Poisson distribution, the extended diagonal model and a maximum number of change-points of 50. To compensate for TADs of different sizes, the selected regions are resized prior to averaging the contact maps. These regions are comprised of the TAD itself and a flanking region of half its size. We calculated the insulation score as described previously⁵⁹. The insulation score was computed using GENOVA's implementation with a rolling window size of 25 kb. The insulation score was then aligned to TAD borders to create heatmaps.

For Extended Data Fig. 6, the compartment-score was calculated as described previously⁶⁰. In short, the compartment score is computed per chromosome arm by obtaining the first eigenvector of the observed over expected (O/E) matrix minus 1. Then this eigenvector is multiplied by the square root of its eigenvalue to obtain the compartment score. To correctly orient the scores so that positive values correspond to compartment A regions, we used the compartment score's correlation to H3K4me1 peaks in wild type cells (Haarhuis et al, manuscript in preparation).

For Extended Data Fig. 6e, we compared the effect of CTCF Y226A F228A on genome organization to that of CTCF and cohesin degradation¹². Raw Hi-C data from GEO accession GSE102884 was converted to HiC-Pro format and ICE normalized. Relative contact probability profiles were generated using GENOVA.

ChIPseq

Samples for ChIPseq samples were prepared and sequenced as previously described¹⁰, with minor changes. The DNA was sheared using Biorupter Pico (Diagenode), 5 cycles 15 seconds on 90 seconds off. Reads were first trimmed using TrimGalore v0.6.0⁶¹, then mapped to hg19 using Bowtie2 v2.3.4⁶² with default settings. Bigwig files were generated with DeepTools v3.1.3⁶³ with the following settings: minimum mapping quality of 15, bin length of 10bp, extending reads to 200bp and reads per kilobase per million reads (RPKM) normalization.

Peaks were called for all samples using MACS2 v2.1.1⁶⁴ with default options. Overlaps between the sets of identified peaks across samples were obtained using BEDtools v2.25.0⁶⁵. Heatmaps were generated using DeepTools⁶³ for the different sets of peaks identified in the wild type cell line, excluding those overlapping blacklisted regions of the genome (ENCODE project consortium, 2012).

ChIP-qPCR

ChIP-qPCR analysis was performed to assess SCC1 and CTCF abundance at specific genomic loci. SCC1 ChIP was performed three times, and on each ChIP three qPCRs were performed in duplicate. A representative qPCR analysis of each ChIP was used for quantification. For CTCF and IgG, two ChIPs were performed in duplicate. Reactions were performed using SYBR No-Rox Mix 2X (BIOLINE) and run on a LightCycler 480 II (Roche). Ct values were determined for input and ChIP samples and subsequently the dCt was converted into a percentage of input. The primers are listed in Extended Data Table 3a.

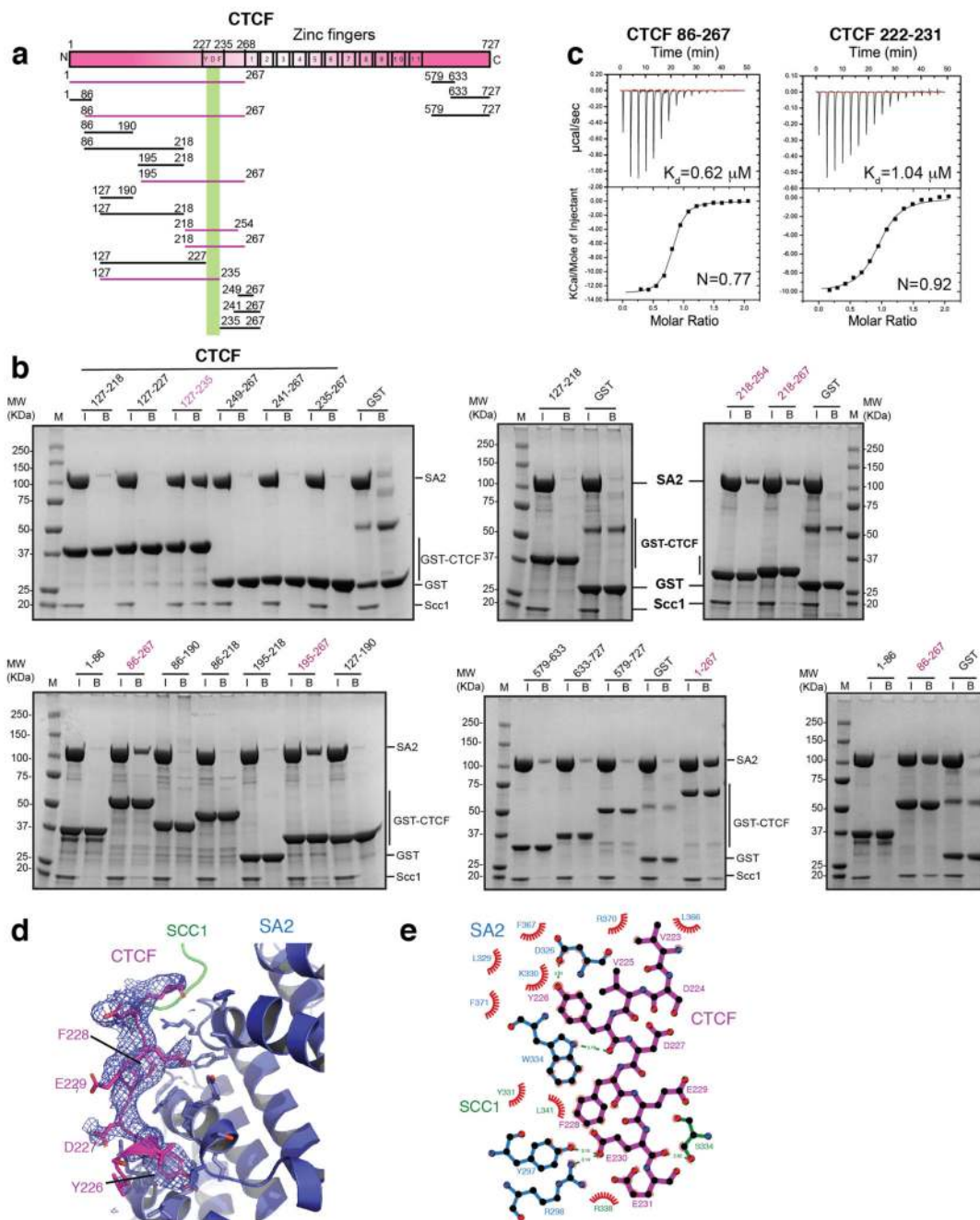
CTCF-associated loops

CTCF sites shown in Hi-C contact matrices were obtained from Haarhuis et al.¹⁰. Briefly, these were generated by intersecting their CTCF peaks with CTCF motifs from JASPAR CORE 2014⁶⁶ using FIMO⁶⁷ to annotate their motif orientation.

RNAseq

Samples for RNAseq were prepared and sequenced as previously described¹⁰. Reads were aligned to hg19 using TopHat v2.1.1⁶⁸ and later counted with HTSeq v0.11.1⁶⁹ using Gencode v19 gene-build as reference. Differentially expressed genes were identified with DESeq2 v1.18.135⁷⁰, with an adjusted p-value threshold of 0.05 and considering only protein coding genes.

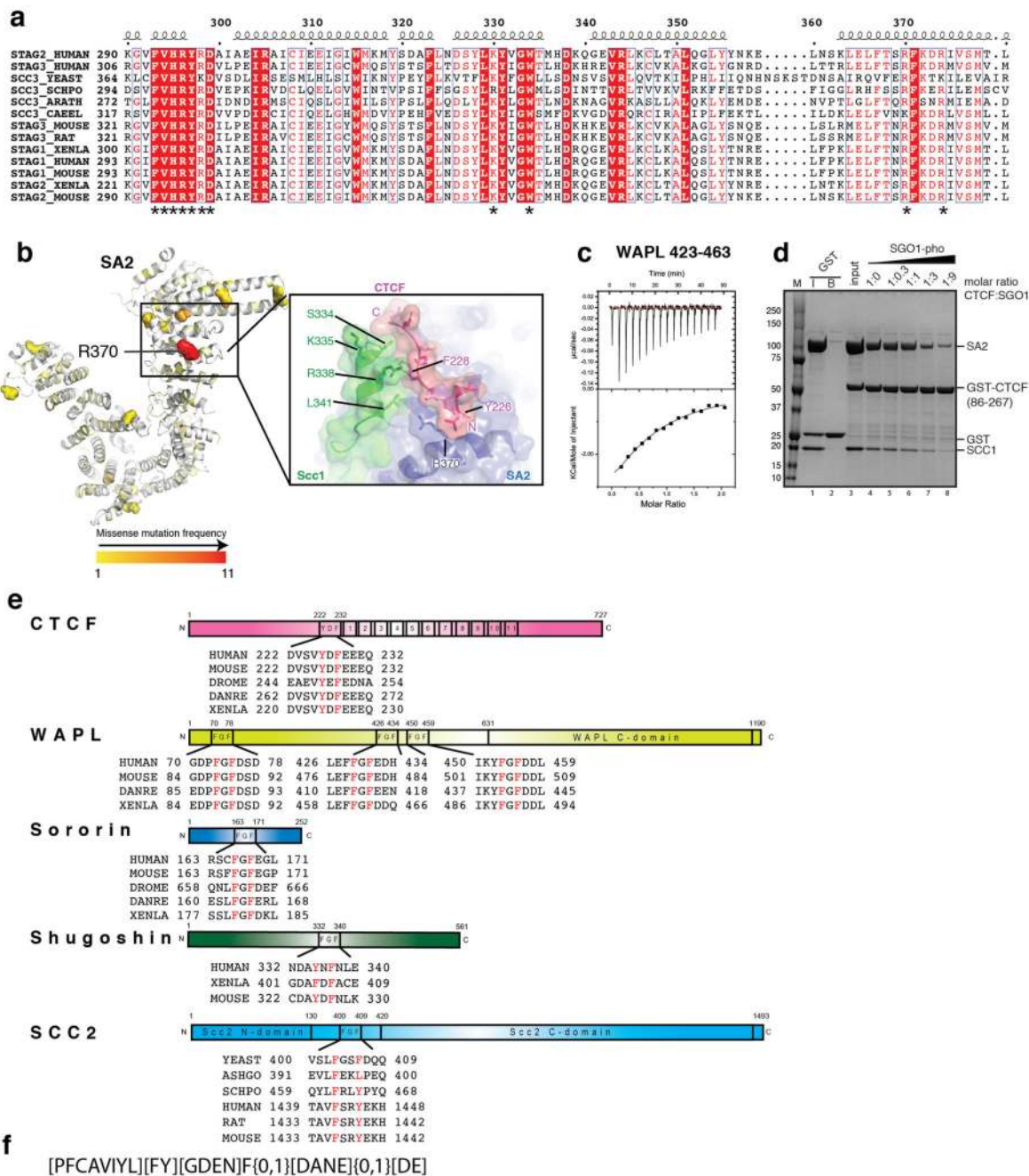
Extended Data



Extended Data Figure 1. Biochemical analysis of CTCF binding to SA2-SCC1.

a, Domain architecture of CTCF. CTCF fragments tested for SA2-SCC1 binding by GST pull-down analysis are indicated. The region retaining SA2-SCC1 is highlighted in magenta. **b**, Summary data showing results of GST pull-downs. The input (I) and the bound (B) fractions were analysed by SDS-PAGE. CTCF fragments that bind SA2-SCC1 are shown in magenta. The experiment was repeated once. **c**, ITC curves. The binding stoichiometry (N) and dissociation constants K_d are indicated. The experiment was repeated three times with consistency. **d**, Fo-Fc omit electron density Fourier map contoured at 3 Sigma. **e**, LIGPLOT

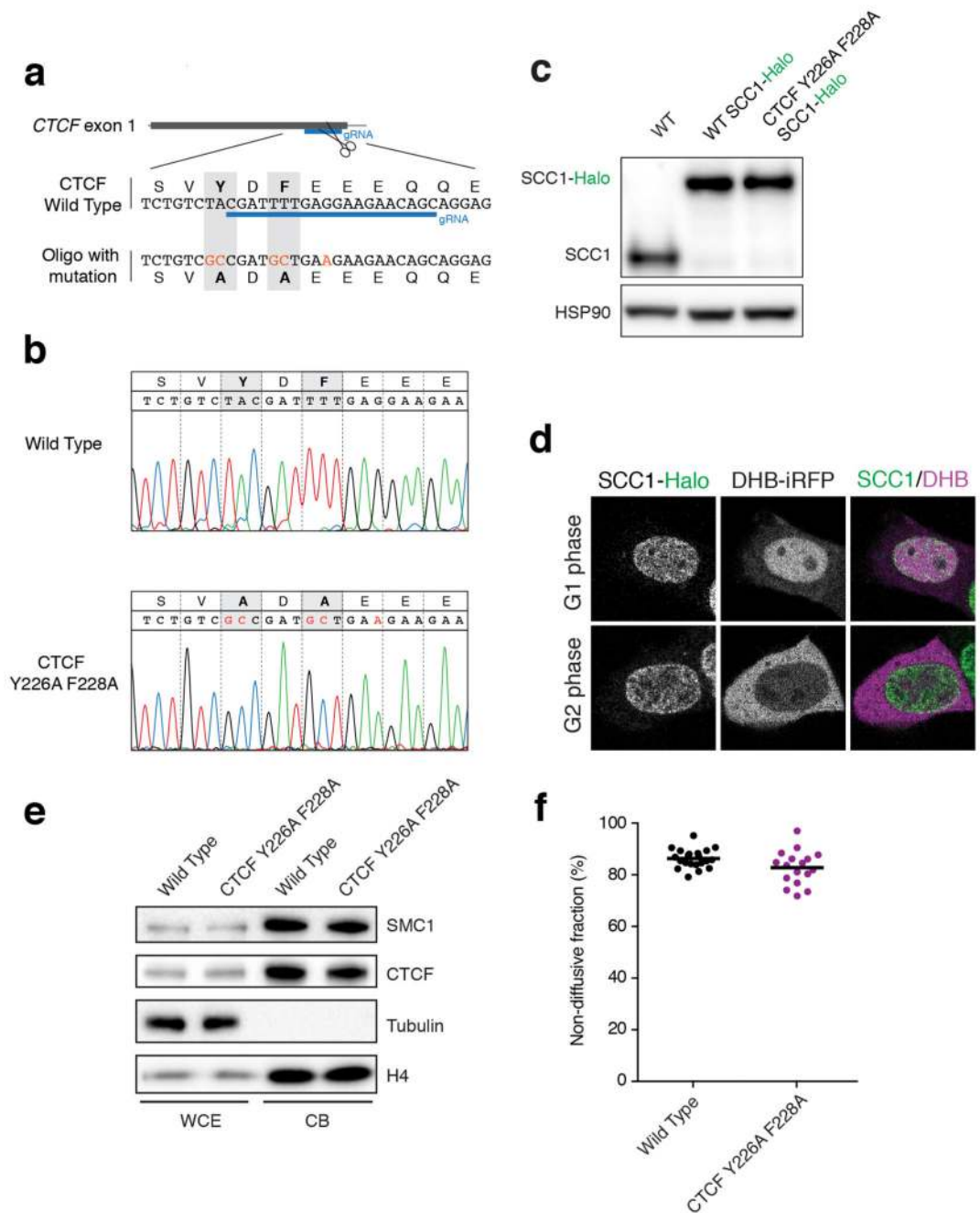
representation of the interaction between the CTCF peptide and SA2-SCC1. The CTCF peptide is shown in magenta, SA2 in blue and SCC1 in green bonds.



Extended Data Figure 2. Analysis of the SA2-SCC1-CTCF structure.

a, Multiple sequence alignment of SA2 (STAG2) orthologs and paralogs. The key amino acid residues engaging CTCF are indicated by (*). **b**, Missense mutation frequencies plotted onto the SA2 structure. R370, a hotspot in SA2, is indicated. The inset shows an overview of mutation hotspots R370 (SA2), Y226, F228 (CTCF) and S334, K335, R338, L341 (SCC1). **c**, ITC progress curves of binding between WAPL 423-463 and SA2-SCC1. **d**, Competition between SGO1 and CTCF for SA2-SCC1 binding. SA2-SCC1 was incubated with GST-CTCF 86-267. Increasing amounts (lanes 4-8; molar ratios are indicated) of T346-

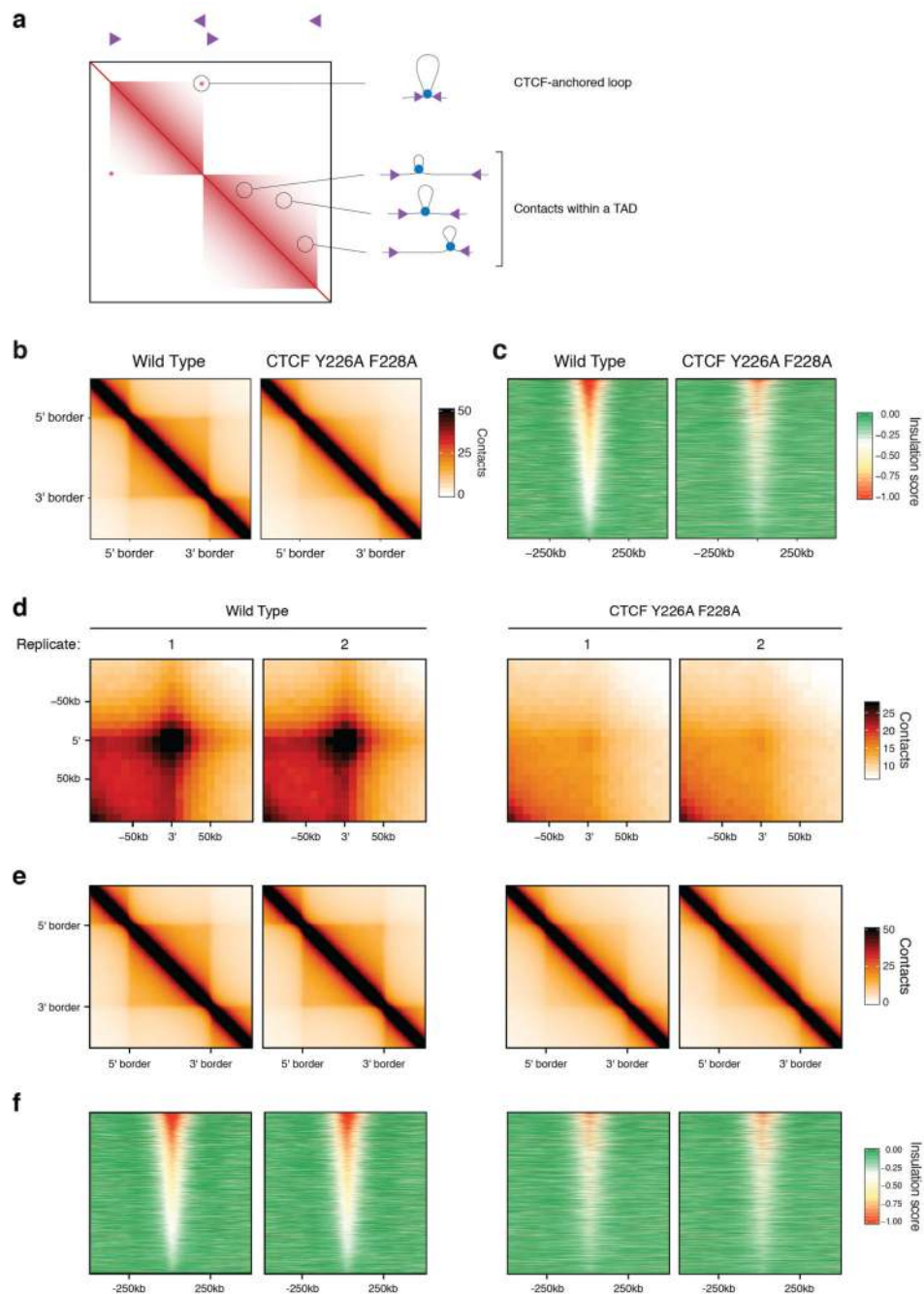
phosphorylated SGO1 peptide spanning 331-349 were added and the input (I) and the bound (B) fraction analysed by SDS-PAGE. The experiment was repeated twice. One representative example is shown. **c**, Domain architecture and sequence alignments of cohesin regulators containing F/YxF motifs. Putative CES interacting residues are highlighted in red. **d**, Regular expression motif used to query the human and yeast proteomes for F/YxF-containing factors. Regular expression syntax: letters denote a specific amino acid; square brackets denote a subset of allowed amino acids; curly brackets denote length variability.



Extended Data Figure 3. Generation of CTCF Y226A F228A cells.

a, Schematic depiction of CRISPR-Cas9 based generation of CTCF Y226A F228A cells. The guide targets cleavage of exon 1 of the *CTCF* gene. The repair oligo renders the gene non-cleavable by Cas9, and simultaneously introduces mutations in the codons encoding Y226 and F228. **b**, The CTCF Y226A F228A mutation was confirmed by Sanger sequencing, including a silent mutation at position 229. **c**, Western blot depicting Halo-tagged SCC1 in wild type and CTCF Y226A F228A cells. The parental wild type cells are included as a control. This experiment was performed once. **d**, Representative images of

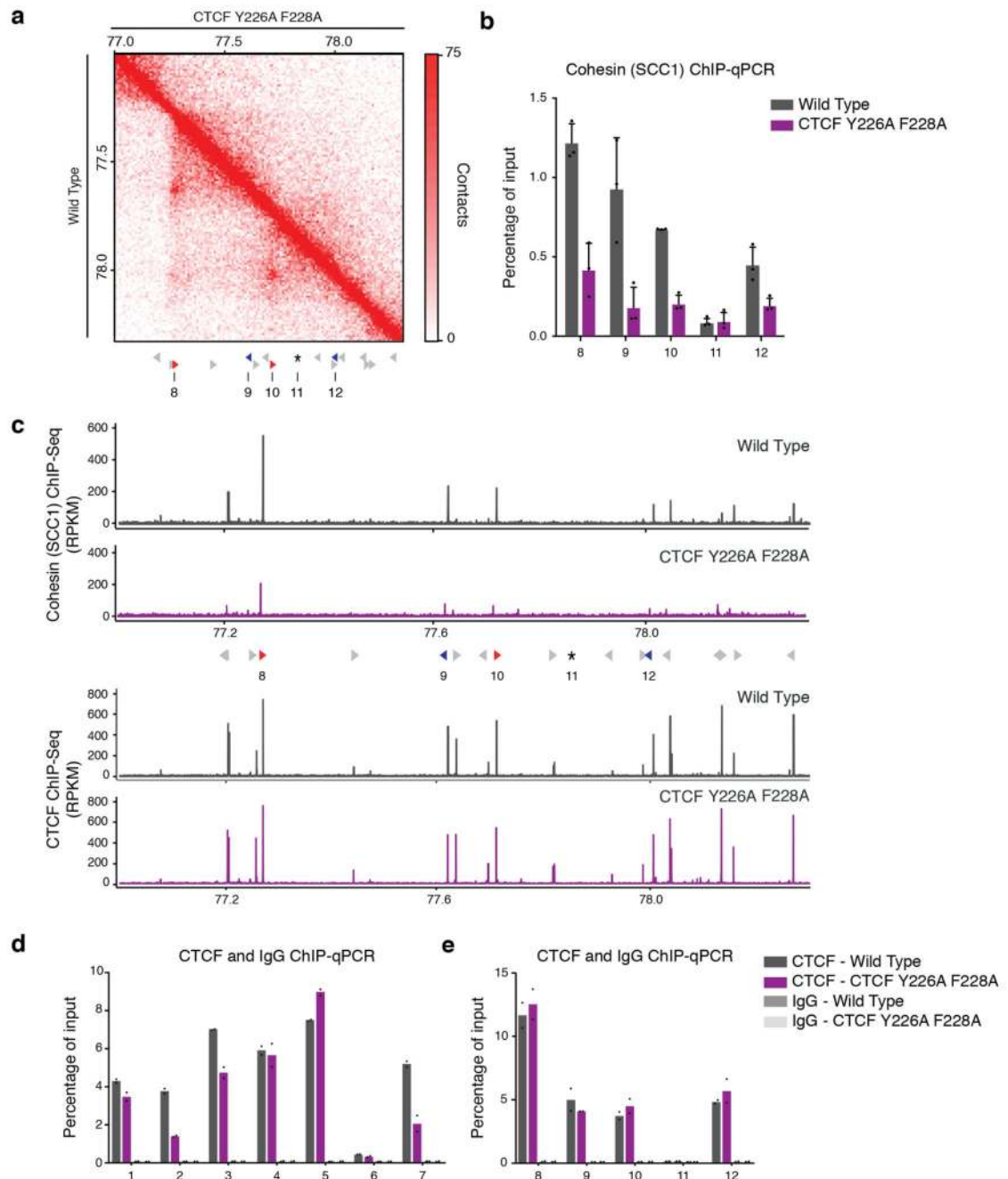
cells in G1 and G2, as indicated by their nuclear and cytoplasmic localization of DHB-iRFP respectively. **e**, Chromatin bound levels of CTCF and SMC1 analyzed by Western Blot. Histone H4 is used as a control for the chromatin fraction. CTCF Y226A F228A mutation does not evidently affect overall CTCF and cohesin levels on chromatin. WCE: whole-cell extract, CB: chromatin bound fraction. This experiment was performed twice with similar results. **f**, Relative SCC1-Halo fluorescence intensity quantified in the unbleached area directly after photobleaching, as a proxy for the chromatin-bound fraction of SCC1. This non-diffusive fraction is not evidently affected by the CTCF Y226A F228A mutation. Individual cells of three independent experiments are plotted as dots and their mean is indicated (21 wild type cells and 17 CTCF Y226A F228A cells were scored).



Extended Data Figure 4. TAD analyses and Hi-C replicates.

a, Schematic depiction of a Hi-C matrix displaying DNA-DNA contacts across a genomic region that includes two TADs. TADs in general are flanked by inwards-pointing CTCF sites (magenta arrows). Signal close to the diagonal line reflects short-range contacts, and contacts spanning longer distances are found further away from the diagonal. The contacts within a TAD are formed by cohesin complexes (blue circles). Cohesin builds loops that it can enlarge until it encounters CTCF. Some TADs are enriched for contacts between the two CTCF sites that lie at their boundaries. These contacts are referred to as CTCF-anchored

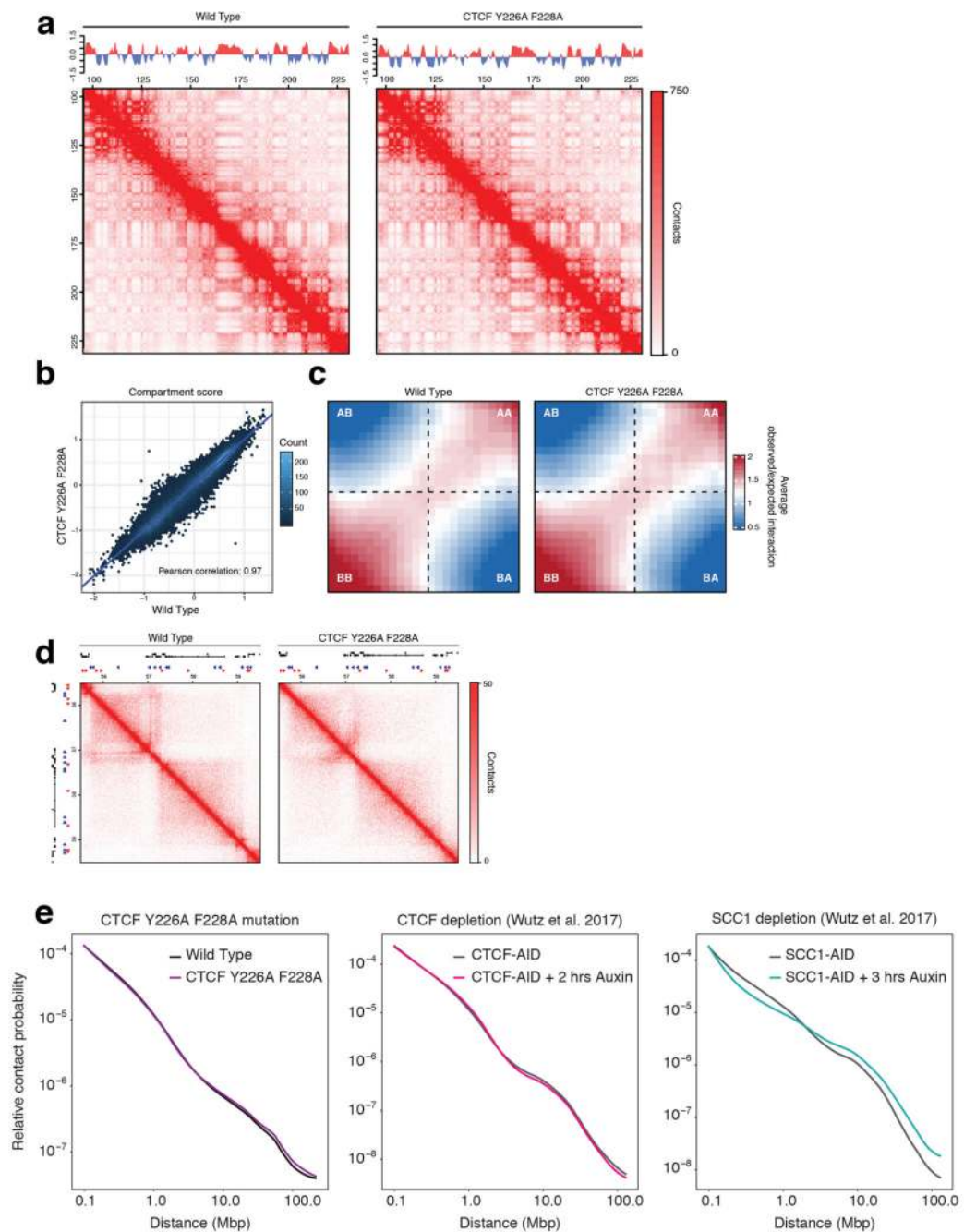
loops. **b**, Aggregate TAD analysis (ATA) depicting the average contact frequency across TADs defined in wild-type cells. **c**, Heatmap of the insulation score⁵⁹ at TAD borders as defined for wild-type cells. **d**, Aggregate peak analysis as in Fig. 3c, using two independent library preps per genotype. **e**, Aggregate TAD analysis for wild-type and CTCF Y226A F228A cells as in (b). **f**, Heatmap of insulation scores at TAD borders for wild-type and CTCF Y226A F228A cells as in (c).



Extended Data Figure 5. CTCF Y226A F228A mutation has little effect on CTCF levels at CTCF sites.

a, Hi-C contact matrix of region chr16: 77000000 - 78300000 at 10 kb resolution for the wild type cell line (lower triangle) and the CTCF Y226A F228A cell line (upper triangle). CTCF sites are depicted below; those selected for qPCR are shown in colour. Red triangles indicate sites with a forward motif and blue triangles indicate sites with a reverse motif. The numbers underneath indicate the qPCR primer pairs shown in (b). Primer pair 11 (indicated with *) is at a locus devoid of SCC1 and CTCF. **b**, ChIP-qPCR analysis of SCC1 (cohesin)

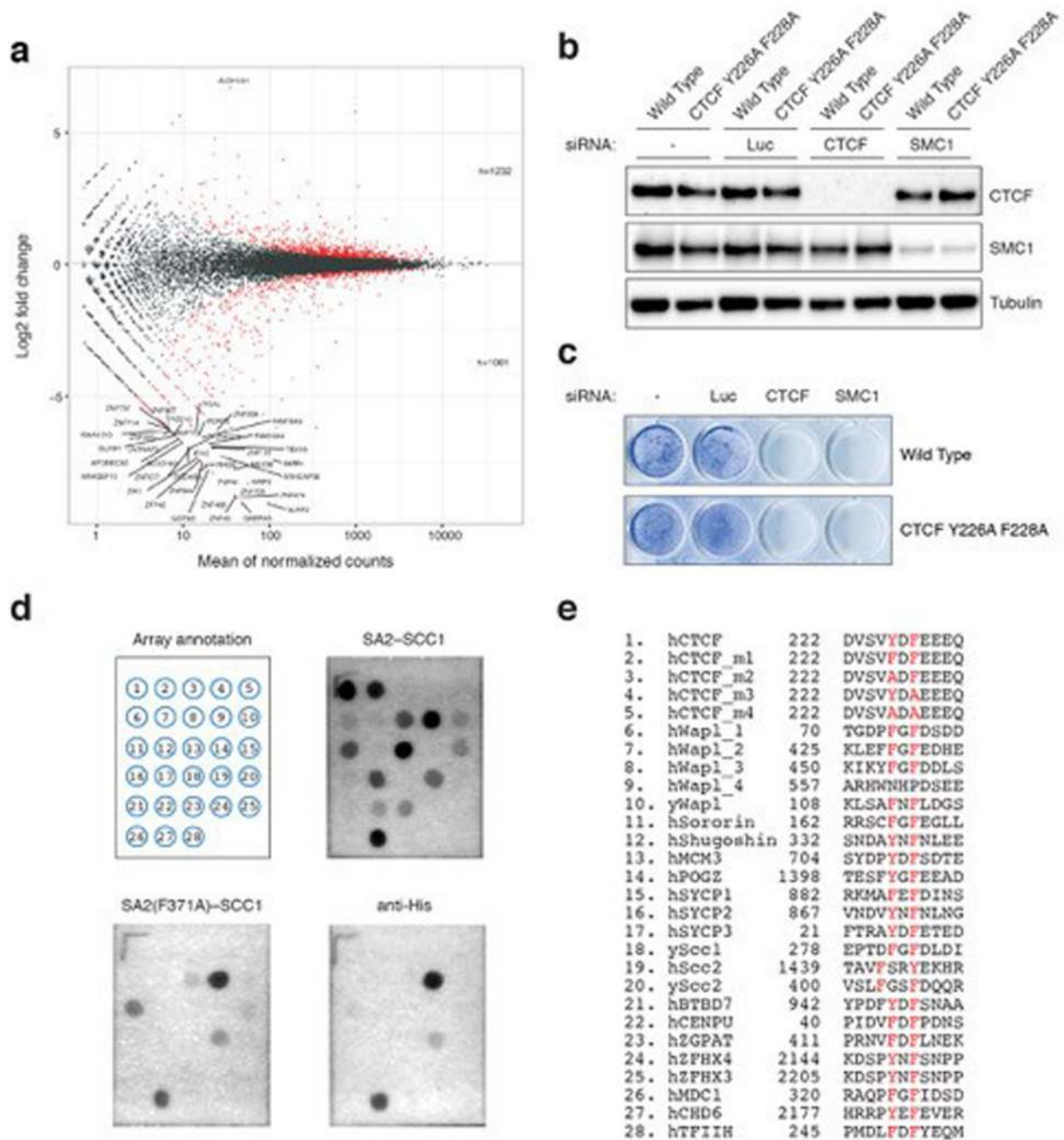
enrichment at the aforementioned CTCF sites and control locus (*) in wild type and CTCF Y226A F228A cells. The mean of three independent ChIP experiments is shown with the standard deviation. **c**, ChIP-seq tracks for SCC1 and CTCF at region chr16: 77000000 - 78300000 in wild type and CTCF Y226A F228A cells. The loci used for ChIP-qPCR analysis are indicated below the SCC1 ChIP-Seq tracks. **d**, ChIP-qPCR analysis of CTCF abundance at loci 1 to 7, as described in Fig. 3d. Analysis includes IgG as a control. The mean of two independent ChIP experiments is shown. Details about replicates are shown in the Methods. **e**, ChIP-qPCR analysis of CTCF abundance at loci 8 to 12, as described in Extended Data Fig. 4a. Analysis includes IgG as a control. The mean of two independent ChIP experiments is shown. Details about replicates are shown in the extended methods.



Extended Data Figure 6. Compartmentalization is largely maintained in cells containing the CTCF Y226A F228A mutation.

a, Hi-C contact matrices of the q-arm of chromosome 2 at 500 kb resolution. The corresponding compartment scores are plotted above. **b**, Genome-wide comparison of compartment scores for wild type and CTCF Y226A F228A cells (Pearson correlation = 0.97). **c**, Saddle plots representing the interaction between A and B compartments. **d**, A region of chromosome 1 (chr1: 55500000 - 59500000) at 10kb resolution that harbours no obvious CTCF-anchored loops. **e**, Relative contact probability profiles for wild type and

CTCF Y226A F228A mutant cells (left), compared to previously published contact profiles upon degradation of CTCF (middle) or SCC1 (right). The CTCF Y226A F228A mutation, like CTCF depletion, only slightly affects the contact probability profile.



Extended Data Figure 7. Identification of CES ligands.

a, MA plot depicting the log₂ fold change in gene expression in relation to the mean of the normalized counts for each gene. Differentially expressed genes (adjusted p-value < 0.05, two-tailed Wald test adjusted for multiple testing using the Benjamini-Hochberg procedure) are shown in red. Gene names are included for the 40 genes with the highest fold-change. **b**, Western blot assessing knockdown of CTCF and the cohesin subunit SMC1 upon transfection with a control siRNA targeting Luciferase (Luc) or siRNAs targeting CTCF or SMC1. This experiment was performed twice with similar results. **c**, Colony formation assay

of wild type and CTCF Y226A F228A cells upon transfection with a control siRNA targeting Luciferase (Luc) or siRNAs targeting CTCF or SMC1. CTCF remains essential for viability in CTCF Y226A F228A cells. This experiment was performed four times with similar results. **d**, Peptide array annotation (top left), binding of SA2-SCC1 (top right) or SA2 F371A-SCC1 mutant (bottom left) and antibody control (bottom right). Three independent experiments were done with consistency. One representative example is shown. **e**, Amino acid sequences of the peptides. Predicted lead-anchoring residues are colored in red.

Extended Data Table 1
Summary of ITC and XRAY data collection and refinement statistics.

Protein	Residues	K_d (μ M)	ΔH (kcal/mol)	$T\Delta S$ (kcal/mol)	ΔG (kcal/mol)	N_{\ddagger}
CTCF*	222-231	1.04 ± 0.20	-11.08 ± 0.70	-2.92 ± 0.82	-8.16 ± 0.09	0.93 ± 0.04
CTCF†	86-267	0.62	-13.16	-4.61	-8.54	0.78
Wapl†	423-463	32.8	-6.66	-0.54	-6.11	0.62
Wapl†	447-462	78.7	-6.81	-1.20	-5.60	1.00
Shugoshin†	331-341	13.5	-10.67	-4.02	-6.64	0.83
Shugoshin†	331-349 (pT346)§	2.32	-20.00	-12.30	-7.69	0.89

SA2-Scc1-CTCF PDB 6QNX	
Data collection	
Space group	P2 ₁ 2 ₁ 2 ₁
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	79.02, 107.25, 176.49
Resolution (Å)	45.81–2.70 [^]
R_{sym} or R_{merge}	6.9 (175)*
$\ \sigma I$	12.0 (0.8)*
$CC\ 1/2$	0.99 (0.33)*
Completeness (%)	99.6 (99.7)*
Redundancy	4.4 (4.3)*
Refinement	
Resolution (Å)	45.81–2.70
$R_{\text{work}} / R_{\text{free}}$	0.25 / 0.27
No. reflections	46759
No. atoms	16487
SA2	15088
Scc1	1235
Ligand	140 _{CTCF}
B -factors (mean; Å ²)	
SA2	133.4
Scc1	111.3

SA2-Scc1-CTCF PDB 6QNX	
Ligand	143.6 _{CTCF}
R.m.s deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.53

Extended Data Table 2 Quantification of peptide arrays.

Three independent experiments were done with consistency. Peptide spot signal intensities were correlated to the K_d of CTCF wild-type thus yielding a semi-quantitative binding assay⁷¹. Data points are indicated as means \pm standard deviation.

Position	Protein	species	Mutation	Uniprot	Sequence	K_d [μ M]
1	CTCF	Human	Wild-type	P49711	222 DVSVYDFEEEEQ	1.04 \pm 0.16*
2	CTCF	Human	Y226F	P49711	222 DVSVFDFEEEEQ	1.60 \pm 0.03†
3	CTCF	Human	Y226A	P49711	222 DVSVADFEEEEQ	n.b.
4	CTCF	Human	F228A	P49711	222 DVSVYDAEEEEQ	n.b.
5	CTCF	Human	Y226/F228A	P49711	222 DVSVADAELEEQ	n.b.
6	WAPL	Human	Wild-type	Q7Z5K2	70 TGDVFGFDSDD	5.90 \pm 0.95†
7	WAPL	Human	Wild-type	Q7Z5K2	425 KLEFVGFEDHE	12.43 \pm 7.72†
8	WAPL	Human	Wild-type	Q7Z5K2	450 KIKYVGFDDL	2.17 \pm 0.16†
9	WAPL	Human	Wild-type	Q7Z5K2	557 ARHWNHPDSEE	ND
10	WAPL	Yeast	Wild-type	Q99359	108 KLSA ^F NFLDGS	10.65 \pm 3.83†
11	CDCA5	Human	Wild-type	Q96FF9	162 RRSCVGFEGLL	ND
12	SGO1	Human	Wild-type	Q5FBB7	332 SNDA ^Y NFNLEE	n.b.
13	MCM3	Human	Wild-type	P25205	704 SYDPYDFSDTE	1.02 \pm 0.05†
14	POGZ	Human	Wild-type	Q7Z3K3	1398 TESVYGFEEAD	n.b.
15	SYCP1	Human	Wild-type	Q15431	882 RKMA ^F FFDINS	6.71 \pm 3.18†
16	SYCP2	Human	Wild-type	Q9BX26	867 VNDV ^Y NFNLNG	n.b.
17	SYCP3	Human	Wild-type	Q8IZU3	21 FTRAYDFETED	1.96 \pm 0.26†
18	SCC1	Yeast	Wild-type	Q12158	278 EPTDFGFDLDI	n.b.
19	SCC2	Human	Wild-type	Q6KC79	1439 TAVFSRYEKHR	ND
20	SCC2	Yeast	Wild-type	Q04002	400 VSL ^F GSFDQQR	n.b.
21	BTBD7	Human	Wild-type	Q9P203	942 YPDFYDFSNA	n.b.
22	CENPU	Human	Wild-type	Q71F23	40 PIDV ^F DFPDNS	4.03 \pm 0.22†
23	ZGPAT	Human	Wild-type	Q8N5A5	411 PRNV ^F DFLNEK	4.03 \pm 1.19†
24	ZFHX4	Human	Wild-type	Q86UP3	2144 KDSP ^Y NFSNPP	n.b.
25	ZFHX3	Human	Wild-type	Q15911	2205 KDSP ^Y NFSNPP	n.b.
26	MDC1	Human	Wild-type	Q14676	320 RAQP ^F GFIDSD	n.b.
27	CHD6	Human	Wild-type	Q8TD26	2177 HRRP ^Y EFEEVER	ND

Position	Protein	species	Mutation	Uniprot	Sequence	K _d [μM]
28	TFIIH	Human	Wild-type	Q13888	245 PMDLFDFYEQM	n.b.

Extended Data Table 3 Primers and Hi-C statistics

a, Primers used. **b**, Hi-C statistics for replicate library preps. Libraries 1 and 2 are independent preparations, and 2.2 is a deeper re-sequencing of sample 2.1. The independent libraries (1 and 2.1) were used for Extended Data Fig. 4. A merge of replicates 1, 2.1 and 2.2 of the wild type cells, and a merge of replicates 1, 2.1 and 2.2 of CTCF Y226A F228A mutant cells was used for Fig. 3, Fig. 4a and Extended Data Fig. 5a and 6.

c, Hi-C statistics after merging replicates of wild type and CTCF Y226A F228A libraries.

Primerset	Primer orientation	Sequence
Primerpair 1	Forward	GGCACTACAGGACCACGTTT
	Reverse	CCCAATTGTGTCTGCCTTTT
Primerpair 2	Forward	GTGGTGTGGGAAGAGTGTT
	Reverse	GTCAGCTAAACGCCAGGTA
Primerpair 3	Forward	CAAGTTTTCCACCCGCTTTA
	Reverse	GAGCCCTAACACCACTCCAC
Primerpair 4	Forward	GGCTTGGAACTGTTGGTCAT
	Reverse	AGATGGCAGCAGCTTTTCAT
Primerpair 5	Forward	TGATTGTGTACAACAGCTGCAA
	Reverse	ATTTTTAGGTGCCTCGCAGT
Primerpair 6	Forward	CTGAGCCTCCTGCAAAAGTT
	Reverse	CTCTTCTTCGCTCCAGCACT
Primerpair 7	Forward	ACTGCAGCCTCAGCTACCTC
	Reverse	TTTATTGGCATTGCCTCCTC
Primerpair 8	Forward	CAGTCCTTGTGGCTCCTAGC
	Reverse	TCTGGTGTGCCCTGAACATA
Primerpair 9	Forward	CACCTTGTGGACAGTGGTTG
	Reverse	AGCCTGTGAAACAGGGTGAG
Primerpair 10	Forward	TACACGGGTGGCTAAAGGAG
	Reverse	AGCCAGCCAGATGTCAAACCT
Primerpair 11	Forward	CATGCCAGCCAATTATTTT
	Reverse	CTCTCTCCACTTCCCCATT
Primerpair 12	Forward	CACTTTTCCGACCCAGAAGA
	Reverse	GGCCTGGAGAACTCAAACCTG

Genotype	Replicate	Total Pairs	Valid Pairs	Cis	Cis %	Cis < 20kb	Cis > 20kb	Cis ratio
Wild type	1	61118122	60166198	47100811	78,28	7085049	40015762	5,65

Genotype	Replicate	Total Pairs	Valid Pairs	Cis	Cis%	Cis < 20kb	Cis > 20kb	Cis ratio
Wild type	2.1	62631817	61755440	48127333	77,93	7114243	41043090	5,76
Wild type	2.2	190892790	152708260	122528381	80,24	18087008	104441373	5,77
CTCF Y226A F228A	1	63339779	62197640	47164621	75,83	72900092	39874529	5,47
CTCF Y226A F228A	2.1	62326840	61227593	46569997	76,06	7419962	39150035	5,28
CTCF Y226A F228A	2.2	148586127	118816672	93071165	78,33	14814014	78257151	5,28

Genotype	Replicate	Total Pairs	Valid Pairs	Cis	Cis%	Cis < 20kb	Cis > 20kb	Cis ratio
Wild type	1+2.1+2.2	312814428	270823907	217755919	80,40	32286097	185469852	5,74
CTCF Y226A F228A	1+2.1+2.2	272011360	238342505	186805272	78,38	29523837	157281435	5,33

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was funded by EMBL. J.H.I.H, A.S.C. and B.D.R were supported by an ERC CoG (772471 'CohesinLooping'), M.S.v.R. by the Boehringer Ingelheim Fonds, and H.T. and E.d.W. by an ERC StG (637597 'HAP-PHEN') and are part of Oncode Institute which is partly financed by the Dutch Cancer Society. We thank the staff at the ESRF beamline Massif-1. We thank Toby Gibson (EMBL) for advice concerning short linear motifs. We thank James Rhodes and Kim Nasmyth for reagents and advice on Halo-tagging, Robin van der Weide for advice and bioinformatic analyses, and Ron Kerkhoven and the NKI Genomics Core Facility for sequencing.

References

1. Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. Annual review of genomics and human genetics. 2016; 17:17–43.
2. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell. 2016; 164:1110–1121. [PubMed: 26967279]
3. Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. Nature reviews Genetics. 2018; 19:789–800.
4. Yatskevich S, Rhodes J, Nasmyth K. Organization of Chromosomal DNA by SMC Complexes. Annu Rev Genet. 2019
5. Alipour E, Marko JF. Self-organization of domain structures by DNA-loop-extruding enzymes. Nucleic Acids Res. 2012; 40:11202–11212. [PubMed: 23074191]
6. Fudenberg G, et al. Formation of Chromosomal Domains by Loop Extrusion. Cell reports. 2016; 15:2038–2049. [PubMed: 27210764]
7. Sanborn AL, et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A. 2015; 112:E6456–6465. [PubMed: 26499245]
8. Haarhuis JHI, et al. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. Cell. 2017; 169:693–707 e614. [PubMed: 28475897]

9. Nora EP, et al. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. 2017; 169:930–944 e922. [PubMed: 28525758]
10. Wutz G, et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J*. 2017; 36:3573–3599. [PubMed: 29217591]
11. Guo Y, et al. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*. 2015; 162:900–910. [PubMed: 26276636]
12. de Wit E, et al. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*. 2015; 60:676–684. [PubMed: 26527277]
13. Rao SS, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014; 159:1665–1680. [PubMed: 25497547]
14. Vietri Rudan M, et al. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell reports*. 2015; 10:1297–1309. [PubMed: 25732821]
15. Rao SSP, et al. Cohesin Loss Eliminates All Loop Domains. *Cell*. 2017; 171:305–320 e324. [PubMed: 28985562]
16. Gassler J, et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J*. 2017; 36:3600–3618. [PubMed: 29217590]
17. Rubio ED, et al. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A*. 2008; 105:8309–8314. [PubMed: 18550811]
18. Xiao T, Wallace J, Felsenfeld G. Specific sites in the C terminus of CTCF interact with the SA2 subunit of the cohesin complex and are required for cohesin-dependent insulation activity. *Mol Cell Biol*. 2011; 31:2174–2183. [PubMed: 21444719]
19. Hara K, et al. Structure of cohesin subcomplex pinpoints direct shugosin-Wapl antagonism in centromeric cohesion. *Nature Structural & Molecular Biology*. 2014
20. Pezzi N, et al. STAG3, a novel gene encoding a protein involved in meiotic chromosome pairing and location of STAG3-related genes flanking the Williams-Beuren syndrome deletion. *FASEB J*. 2000; 14:581–592. [PubMed: 10698974]
21. Orgil O, et al. A conserved domain in the scc3 subunit of cohesin mediates the interaction with both med1 and the cohesin loader complex. *PLoS genetics*. 2015; 11:e1005036. [PubMed: 25748820]
22. Roig MB, et al. Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Lett*. 2014; 588:3692–3702. [PubMed: 25171859]
23. Forbes SA, et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res*. 2017; 45:D777–D783. [PubMed: 27899578]
24. Beckouet F, et al. Releasing Activity Disengages Cohesin's Smc3/Sec1 Interface in a Process Blocked by Acetylation. *Mol Cell*. 2016; 61:563–574. [PubMed: 26895425]
25. Gandhi R, Gillespie PJ, Hirano T. Human Wapl is a cohesin-binding protein that promotes sister-chromatid resolution in mitotic prophase. *Curr Biol*. 2006; 16:2406–2417. [PubMed: 17112726]
26. Kueng S, et al. Wapl controls the dynamic association of cohesin with chromatin. *Cell*. 2006; 127:955–967. [PubMed: 17113138]
27. Liu H, Rankin S, Yu H. Phosphorylation-enabled binding of SGO1-PP2A to cohesin protects sororin and centromeric cohesion during mitosis. *Nat Cell Biol*. 2013; 15:40–49. [PubMed: 23242214]
28. Shintomi K, Hirano T. Releasing cohesin from chromosome arms in early mitosis: opposing actions of Wapl-Pds5 and Sgo1. *Genes Dev*. 2009; 23:2224–2236. [PubMed: 19696148]
29. Ouyang Z, et al. Structure of the human cohesin inhibitor Wapl. *Proceedings of the National Academy of Sciences*. 2013; 110:11355–11360.
30. Krystkowiak I, Davey NE. SLiMSearch: a framework for proteome-wide discovery and annotation of functional modules in intrinsically disordered regions. *Nucleic Acids Res*. 2017; 45:W464–W469. [PubMed: 28387819]
31. Flavahan WA, et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature*. 2016; 529:110–114. [PubMed: 26700815]

32. Hnisz D, et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*. 2016; 351:1454–1458. [PubMed: 26940867]
33. Ouyang Z, Zheng G, Tomchick DR, Luo X, Yu H. Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics. *Mol Cell*. 2016; 62:248–259. [PubMed: 26971492]
34. Chan KL, et al. Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell*. 2012; 150:961–974. [PubMed: 22901742]
35. Buheitel J, Stemmann O. Prophase pathway-dependent removal of cohesin from human chromosomes requires opening of the Smc3–Scc1 gate. *The EMBO Journal*. 2013; 32:666–676. [PubMed: 23361318]
36. Eichinger CS, Kurze A, Oliveira RA, Nasmyth K. Disengaging the Smc3/kleisin interface releases cohesin from Drosophila chromosomes during interphase and mitosis. *EMBO J*. 2013; 32:656–665. [PubMed: 23340528]
37. Sedeno Cacciatore A, Rowland BD. Loop formation by SMC complexes: turning heads, bending elbows, and fixed anchors. *Curr Opin Genet Dev*. 2019; 55:11–18. [PubMed: 31108424]
38. Tang Z, et al. CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell*. 2015; 163:1611–1627. [PubMed: 26686651]
39. Nagy G, et al. Motif oriented high-resolution analysis of ChIP-seq data reveals the topological order of CTCF and cohesin proteins on DNA. *BMC Genomics*. 2016; 17:637. [PubMed: 27526722]
40. Kschonsak M, et al. Structural Basis for a Safety-Belt Mechanism That Anchors Condensin to Chromosomes. *Cell*. 2017; 171:588–600 e524. [PubMed: 28988770]
41. Ganji M, et al. Real-time imaging of DNA loop extrusion by condensin. *Science*. 2018; 360:102–105. [PubMed: 29472443]
42. Li Y, et al. Structural basis for Scc3-dependent cohesin recruitment to chromatin. *eLife*. 2018; 7
43. Studier FW. Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*. 2005; 41:207–234. [PubMed: 15915565]
44. Bowler MW, et al. MASSIF-1: a beamline dedicated to the fully automatic characterization and data collection from crystals of biological macromolecules. *Journal of synchrotron radiation*. 2015; 22:1540–1547. [PubMed: 26524320]
45. Svensson O, Malbet-Monaco S, Popov A, Nurizzo D, Bowler MW. Fully automatic characterization and data collection from crystals of biological macromolecules. *Acta crystallographica Section D, Biological crystallography*. 2015; 71:1757–1767. [PubMed: 26249356]
46. Svensson O, Gilski M, Nurizzo D, Bowler MW. Multi-position data collection and dynamic beam sizing: recent improvements to the automatic data-collection algorithms on MASSIF-1. *Acta crystallographica Section D, Structural biology*. 2018; 74:433–440. [PubMed: 29717714]
47. Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta crystallographica Section D, Biological crystallography*. 2010; 66:133–144. [PubMed: 20124693]
48. Winn MD, et al. Overview of the CCP4 suite and current developments. *Acta crystallographica Section D, Biological crystallography*. 2011; 67:235–242. [PubMed: 21460441]
49. McCoy AJ, et al. Phaser crystallographic software. *Journal of applied crystallography*. 2007; 40:658–674. [PubMed: 19461840]
50. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta crystallographica Section D, Biological crystallography*. 2010; 66:486–501. [PubMed: 20383002]
51. Adams PD, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica Section D, Biological crystallography*. 2010; 66:213–221. [PubMed: 20124702]
52. Chen VB, et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta crystallographica Section D, Biological crystallography*. 2010; 66:12–21. [PubMed: 20057044]
53. Yin M, et al. Molecular mechanism of directional CTCF recognition of a diverse range of genomic sites. *Cell Res*. 2017; 27:1365–1377. [PubMed: 29076501]

54. Rhodes JDP, et al. Cohesin Can Remain Associated with Chromosomes during DNA Replication. *Cell reports*. 2017; 20:2749–2755. [PubMed: 28930671]
55. Servant N, et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology*. 2015; 16:259. [PubMed: 26619908]
56. Yang T, et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res*. 2017; 27:1939–1949. [PubMed: 28855260]
57. Durand NC, et al. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. 2016; 3:95–98. [PubMed: 27467249]
58. Imakaev M, et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods*. 2012; 9:999–1003. [PubMed: 22941365]
59. Crane E, et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*. 2015; 523:240–244. [PubMed: 26030525]
60. Flyamer IM, et al. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. 2017; 544:110–114. [PubMed: 28355183]
61. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*. 2011; 17:10.
62. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nature methods*. 2012; 9:357–359. [PubMed: 22388286]
63. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014; 42:W187–191. [PubMed: 24799436]
64. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012; 7:1728–1740. [PubMed: 22936215]
65. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26:841–842. [PubMed: 20110278]
66. Mathelier A, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2014; 42:D142–147. [PubMed: 24194598]
67. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011; 27:1017–1018. [PubMed: 21330290]
68. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*. 2013; 14:R36. [PubMed: 23618408]
69. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015; 31:166–169. [PubMed: 25260700]
70. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*. 2014; 15:550. [PubMed: 25516281]
71. Landgraf C, et al. Protein interaction networks by proteome peptide scanning. *PLoS biology*. 2004; 2:E14. [PubMed: 14737190]

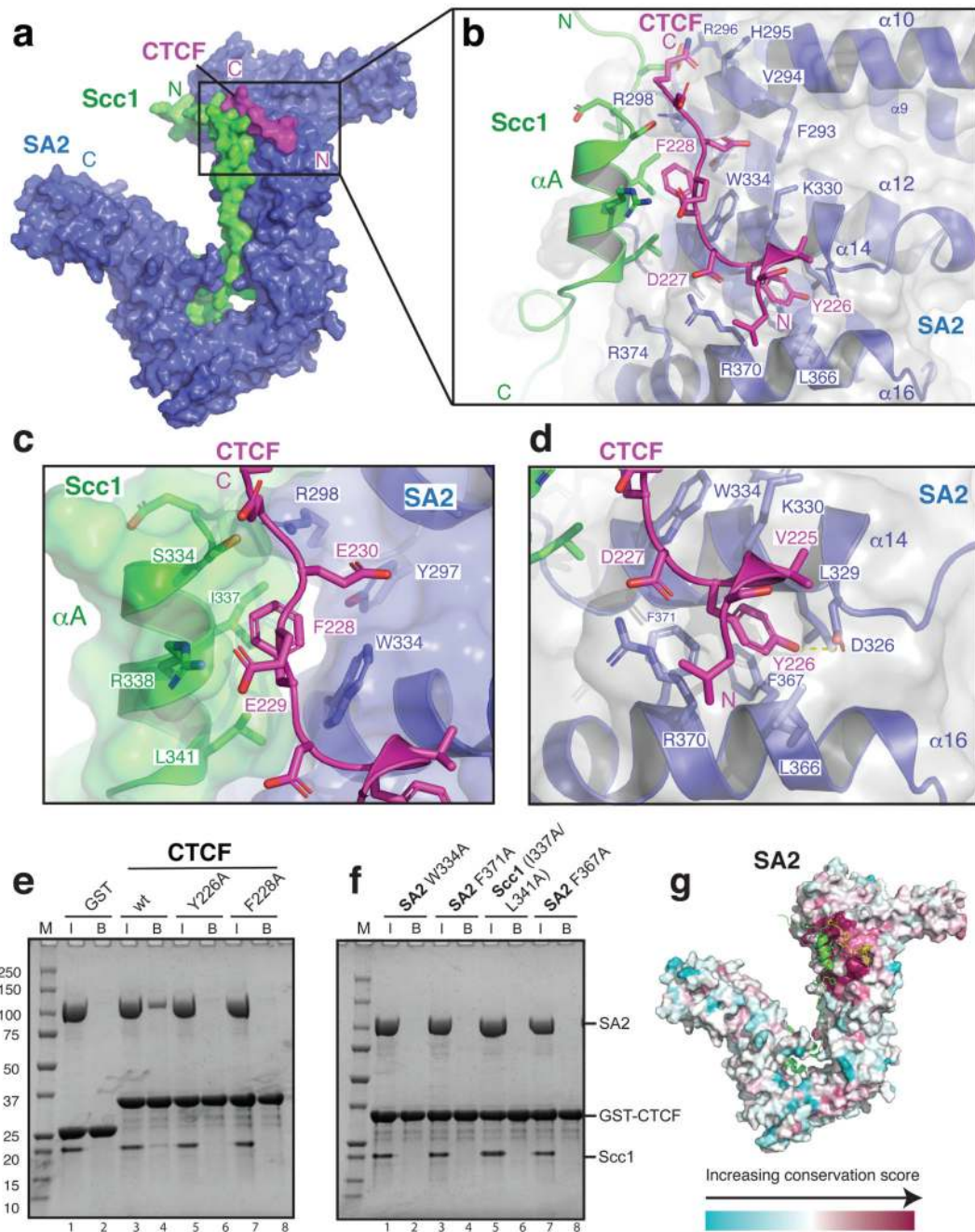


Figure 1. Structure of the SA2-SCC1-CTCF complex.

a, Surface-rendered cartoon of the SA2-SCC1-CTCF complex colored in blue, green and magenta, respectively. **b**, Detailed view of the binding interface with SA2 residues in blue, SCC1 in green and CTCF in magenta. **c**, Details of the composite binding pocket around CTCF F228 and **d**, CTCF Y226. **e**, GST pull-down analysis of CTCF and **f**, SA2 or SCC1 variants. Controls are shown in panel **e** (lanes 1-2). Experiments were done once. **g**, SA2 is surface-rendered and colored according to sequence conservation.

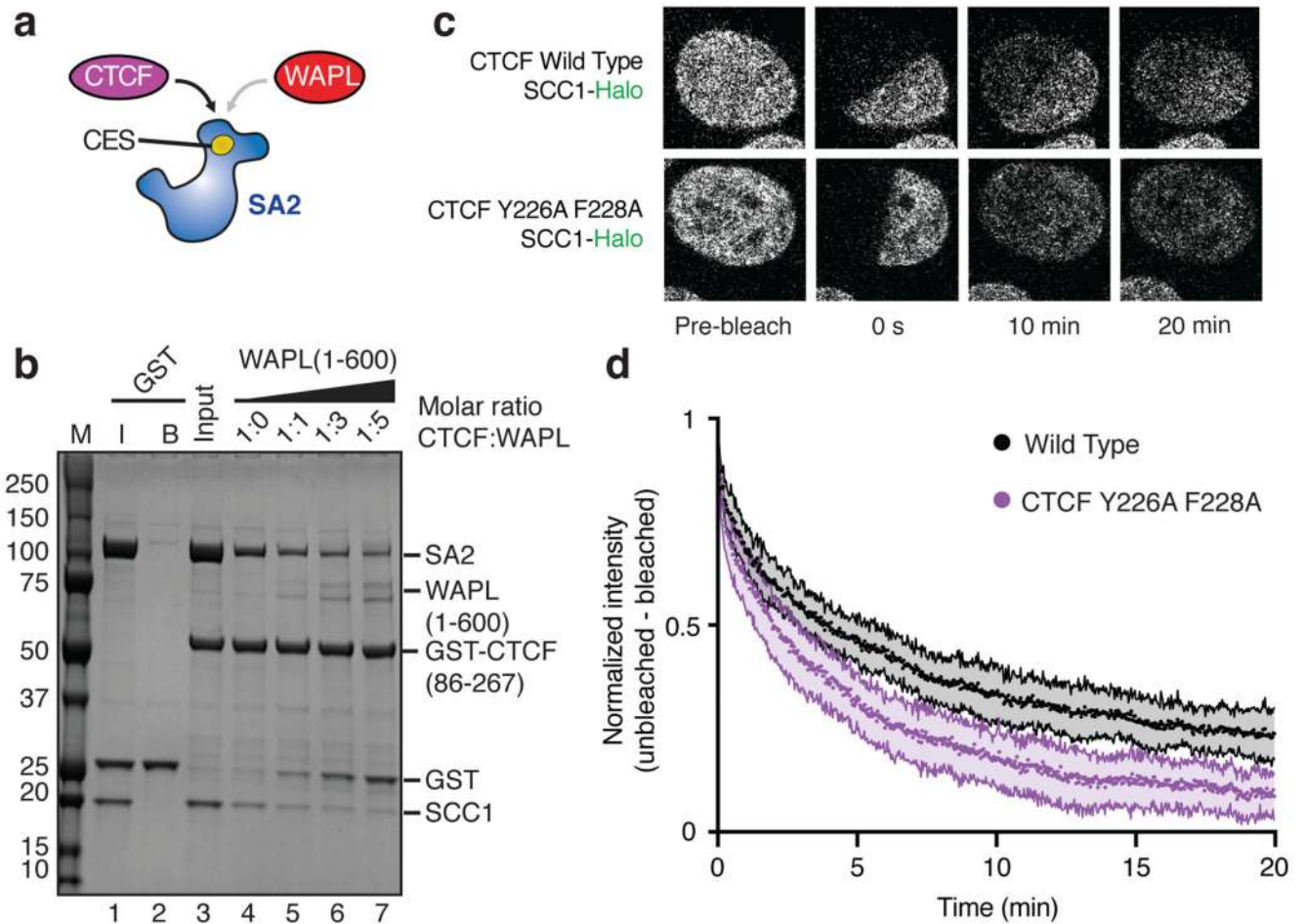


Figure 2. CTCF interaction stabilizes cohesin on DNA.

a. Schematic representation of competition between CTCF and WAPL. **b.** Increasing amounts of WAPL residues 1-600 (lane 4-7; molar ratios are indicated) were incubated with GST-CTCF and SA2-SCC1 and the bound fraction analysed. Three independent experiments were done with consistency. A representative example is shown. **c.** Example images of cells used in **(d)** at the indicated time points after photobleaching. FRAP was performed in G1 cells (see Extended Data Fig. 3d for details). **d.** Quantification of the FRAP experiments. Averages and standard deviations for 21 wild type cells and 17 CTCF Y226A F228A cells, measured over three independent experiments.

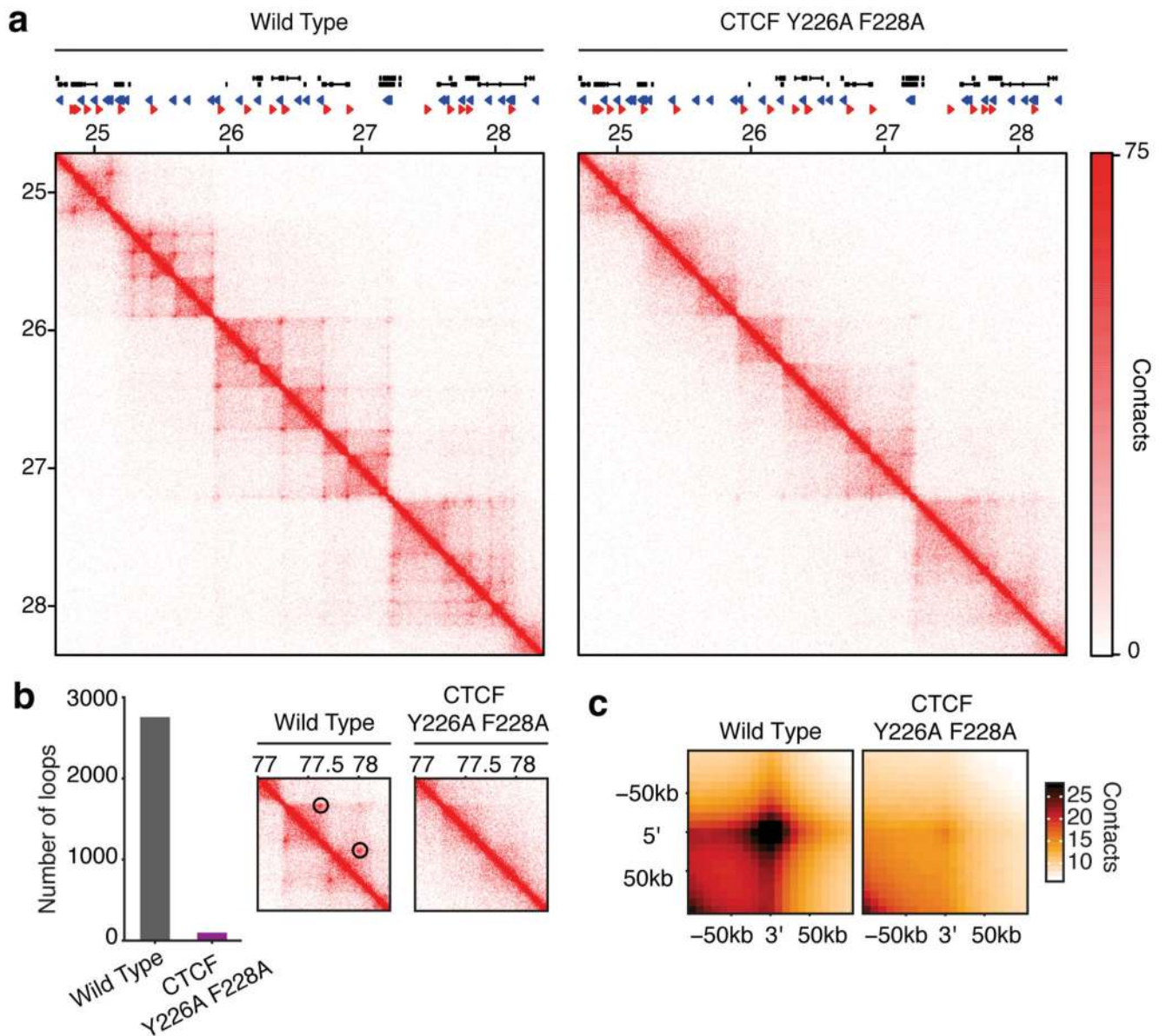


Figure 3. CTCF-CES interaction is required for CTCF-anchored loops.

a, Hi-C contact matrices of the *HoxA* locus at 10 kb resolution, normalized to 100 million contacts per sample. Genes and CTCF sites are depicted above the contact matrices. **b**, Genome-wide quantification of loops using HICCUPS¹⁵. The inset shows an example of called loops for a region of chromosome 16. **c**, Aggregate peak analysis (APA) for the loops defined in wild type cells. The Hi-C signal is averaged across these locations for both cell lines.

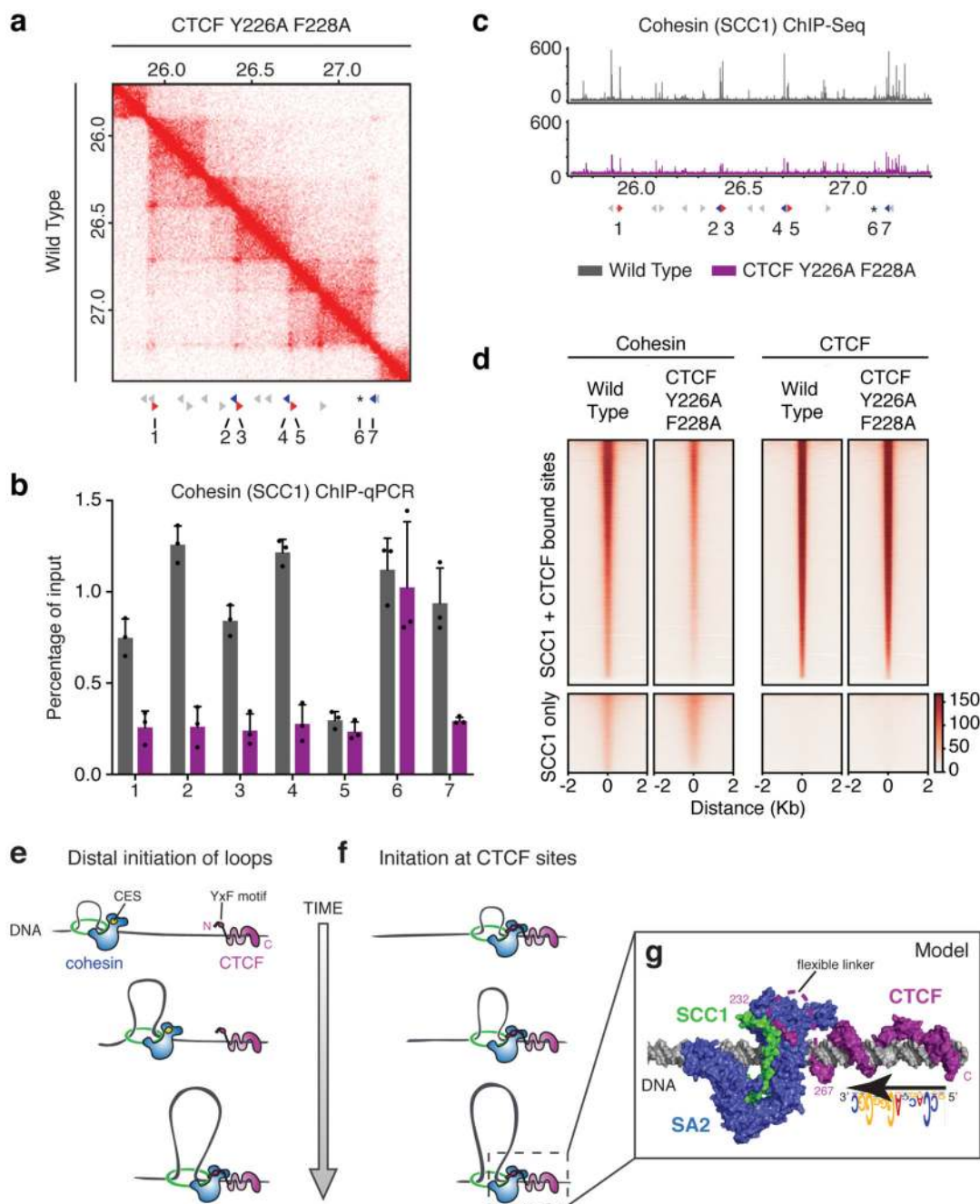


Figure 4. CTCF-CES interaction promotes cohesin localization to CTCF sites.

a, Hi-C contact matrix of a region of chromosome 7 at 10 kb resolution. CTCF sites are depicted below; those selected for qPCR are shown in colour (forward motifs in red, reverse motifs in blue). The numbers underneath indicate the loci used for qPCRs in (b). Locus 6 is the HOTAIRM1 transcription start site (indicated with *). **b**, ChIP-qPCR analysis of SCC1 (cohesin) at the loci depicted in (a). The mean of three independent ChIP experiments is shown with standard deviations. **c**, ChIP-seq tracks for SCC1 of the same region of chromosome 7 as depicted by Hi-C in (a). The ChIP-qPCR loci of (b) are depicted below. **d**,

ChIP-seq heatmap of the cohesin subunit SCC1 (left) and CTCF (right). The depicted sites are selected for being bound in wild type cells by both SCC1 and CTCF (top), or only by SCC1 (bottom). **e**, Cohesin-mediate looping initiates at distal sites until encounter of the N-terminal end of CTCF. **f**, Cohesin-mediated looping starts at CTCF sites. **e**, In the CTCF mutant the YxF motif does not engage the CES on SA2-SCC1 resulting in DNA release. **g**, Molecular model of CTCF and SA2-SCC1 bound to DNA (grey). The YxF motif is separated by a flexible linker spanning residues 232-267 (magenta dotted line) to the C-terminal DNA binding domain of CTCF.