ESCO1 and CTCF enable formation of long chromatin loops by protecting cohesin^{STAG1} from WAPL

Gordana Wutz[1], Rene Ladurner[1†], Brian St. Hilaire[2,3,4], Roman Stocsits[1], Kota Nagasaka[1], Benoit Pignard[1], Adrian Sanborn[2,6], Wen Tang[1], Csilla Várnai[5,12], Miroslav Ivanov[1¶], Stefan Schoenfelder[5,§], Petra van der Lelij[1], Xingfan Huang[2,8,9], Gerhard Dürnberger[7‡], Elisabeth Roitinger[7], Karl Mechtler[1,7], Iain F. Davidson[1], Peter Fraser[5,††], Erez Lieberman Aiden[2,3,4,8,10,11,*] and Jan-Michael Peters[1,*]

1 Research Institute of Molecular Pathology (IMP), Vienna Biocentre (VBC), Campus-Vienna-Biocentre 1, 1030 Vienna, Austria

2 The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA

3 Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

4 Center for Theoretical Biological Physics, Rice University, Houston, TX 77030, USA

5 Nuclear Dynamics Programme, The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, United Kingdom

6 Department of Computer Science, Stanford University, Stanford, CA 94305, USA

7 Institute of Molecular Biotechnology, Vienna Biocentre (VBC), Vienna, Austria

8 Departments of Computer Science and Computational and Applied Mathematics, Rice University, Houston, TX 77030, USA

9 Departments of Computer Science and Genome Sciences, University of Washington, Seattle, WA 98195, USA

10 Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

11 Shanghai Institute for Advanced Immunochemical Studies, Shanghai Tech University, Shanghai, China

12 Centre for Computational Biology, University of Birmingham, B15 2TT Birmingham, UK

† Present address: Department of Biochemistry, Stanford University School of Medicine, Stanford, CA, USA

‡ Present address: Gregor Mendel Institute of Molecular Plant Biology Austrian Academy of Sciences,Dr. Bohr-Gasse 3,1030 Vienna, Austria

¶ Present address: The Francis Crick Institute, London, UK

§ Present address: Epigenetics Programme, The Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, United Kingdom

†† Present address: Department of Biological Science, Florida State University, Tallahassee, FL 32306, USA

Running title: CTCF protects cohesin^{STAG1} from WAPL

Key words: Chromatin structure / cohesin / genome organization / CTCF / TADs and loops

To whom correspondence should be addressed: Tel: E-mail: peters@imp.ac.at and erez@erez.com

Abstract

Eukaryotic genomes are folded into loops. It is thought that these are formed by cohesin complexes *via* extrusion, either until loop expansion is arrested by CTCF or until cohesin is removed from DNA by WAPL. Although WAPL limits cohesin's chromatin residence time to minutes, it has been reported that some loops exist for hours. How these loops can persist is unknown. We show that during G1-phase, mammalian cells contain acetylated cohesin^{STAG1} which binds chromatin for hours, whereas cohesin^{STAG2} binds chromatin for minutes. Our results indicate that CTCF and the acetyltransferase ESCO1 protect a subset of cohesin^{STAG1} complexes from WAPL, thereby enable formation of long and presumably long-lived loops, and that ESCO1, like CTCF, contributes to boundary formation in chromatin looping. Our data are consistent with a model of nested loop extrusion, in which acetylated cohesin^{STAG1} forms stable loops between CTCF sites, demarcating the boundaries of more transient cohesin^{STAG2} extrusion activity.

Introduction

In eukaryotic interphase cells, cohesin complexes are essential for the formation and maintenance of numerous chromatin loops (Gassler et al., 2017; Hadjur et al., 2009; Nativio et al., 2009; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017). These long-range interactions are thought to have both structural and regulatory functions, in the latter case by contributing to recombination and gene regulation (reviewed in Lin et al., 2018; Merkenschlager and Nora, 2016). Hi-C experiments have revealed that chromatin interactions are either enriched in genomic regions called topologically associating domains (TADs) or appear as more pronounced localized interactions which in Hi-C maps are visible as “dots” or on the edge of TADs as “corner peaks” (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014*;* note that in Hi-C maps only these dots and corner peaks are referred to as loops, even though TADs are also formed by looping of chromatin).

Most chromatin loops that are mediated by cohesin are anchored at genomic sites that are bound by the insulator protein CTCF (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014) with which cohesin co-localises genome-wide (Parelho et al., 2008; Wendt et al., 2008). Even though loop anchors can be hundreds of kilobases or even Mega-base pairs (Mb) apart in the linear genome, the CTCF consensus motifs that are found at these sites are typically oriented towards each other, a phenomenon known as “the CTCF convergence rule” (de Wit et al., 2015; Rao et al., 2014; Vietri Rudan et al., 2015). The mechanism by which cohesin and CTCF contribute to the generation and maintenance of loops is unknown, but an attractive hypothesis posits that cohesin acts by extruding loops of genomic DNA until it encounters convergently oriented CTCF sites (Fudenberg et al., 2016; Sanborn et al., 2015). Recent single-molecule experiments have provided evidence that budding yeast condensin, which like cohesin belongs to the family of “structural maintenance of chromosomes” (SMC) family of complexes (Strunnikov et al., 1993), can extrude loops of DNA *in vitro* (Ganji et al., 2018).

Cohesin is a protein complex composed of multiple subunits. Three of these, SMC1, SMC3 and SCC1 (also known as RAD21 and Mcd1) form tri-partite rings, which in replicating cells can entrap newly synthesised DNA molecules to mediate sister chromatid cohesion (Haering et al., 2008). During quiescence (G0) and G1, cohesin is dynamically released from chromatin via the activity of the protein WAPL and has a mean chromatin residence time of 8-25 minutes (Gerlich et al., 2006; Kueng et al., 2006; Tedeschi et al., 2013). Despite cohesin’s dynamic interaction with chromatin, recent experiments suggest that some chromatin loops can persist over a significantly longer timescale of several hours (Vian et al., 2018). How this can occur is not understood. It is possible that chromatin loops are maintained after cohesin has been unloaded, but the observation that experimentally induced degradation of cohesin’s SCC1 subunit leads to disappearance of most TADs and loops within 15 minutes argues against this possibility (Wutz et al., 2017). More plausible scenarios are that long-lived chromatin loops are maintained by multiple short-lived cohesin complexes, or that cohesin can be protected from WAPL so that it can persist on chromatin for longer periods of time.

Precedence for the regulatability of cohesin’s residence times on chromatin comes from studies of proliferating somatic cells and from meiotic cells. During S phase of somatic mammalian cells, around half of all cohesin complexes become protected from WAPL via an incompletely understood mechanism that depends on acetylation of cohesin’s SMC3 subunit (Ben-Shahar et al., 2008; Ünal et al., 2008) by the acetyltransferases ESCO1 and ESCO2 and on the subsequent recruitment to cohesin of the protein sororin (Ladurner et al., 2016; Lafont et al., 2010; Nishiyama et al., 2010). This protection from WAPL increases cohesin’s chromatin residence time to many hours and enables cohesive cohesin complexes to maintain sister chromatid cohesion from S phase until the subsequent mitosis (Gerlich et al., 2006; Schmitz et al., 2007). An even more dramatic prolongation of cohesin’s residence time is thought to exist in mammalian

oocytes. In these cells, a meiotic form of cohesin, cohesin^{REC8} , establishes cohesion during pre-meiotic S phase already before birth and then maintains it, depending on the species, for months or years until meiosis is completed during oocyte maturation cycles after puberty (Burkhardt et al., 2016; Tachibana-Konwalski et al., 2010).

Long residence times of cohesin on chromatin can also be experimentally induced by depletion of WAPL. This leads to a dramatic re-localisation of cohesin from a diffuse nuclear pattern into axial chromosomal domains termed vermicelli, which are thought to represent the base of chromatin loops (Tedeschi et al., 2013). This is accompanied by an increase in the number of long DNA loops and by chromatin compaction (Gassler et al., 2017; Haarhuis et al., 2017; Wutz et al., 2017). These findings imply that the residence time of cohesin on chromatin can be regulated in post-replicative cells in order to maintain cohesion over long periods of time, and they show that prolonging cohesin's residence time experimentally can have major effects on genome organization (however, it is important to note that the stabilisation of cohesive cohesin on chromatin during S and G2 phase does not detectably alter genome architecture, which implies that cohesion and chromatin looping are mediated by distinct populations of cohesin; Holzmann et al., 2019 and references therein). However, to date, a cohesin population with long chromatin residence times has not been identified in G1 phase, where some chromatin loops have been reported to exist for hours (Vian et al., 2018).

In mammalian somatic cells, SCC1 associates with the subunits STAG1 or STAG2 to form two distinct tetrameric cohesin core complexes (Losada et al., 2000; Sumara et al., 2000). Although a recent study reported differences in the chromatin localization patterns for STAG1 and STAG2 and their contributions to chromatin organisation (Kojic et al., 2018), it is unclear if these two forms of cohesin exhibit different residence times and incompletely understood whether they play distinct roles in loop formation. Here, we show that cohesin^{STAG1} and cohesin^{STAG2} do indeed display functional differences, both with respect to their residence times and in their ability to structure chromatin. We find that cohesin^{STAG1} complexes are more highly acetylated, interact more stably with CTCF, form larger chromatin loops, and have a longer residence time on chromatin than cohesin^{STAG2} complexes, consistent with the proposed existence of long-lived chromatin loops. This stabilization of cohesin^{STAG1} depends on CTCF and ESCO1. Depletion of ESCO1 also decreases the insulation between TADs, in a manner similar to that observed following CTCF depletion. Furthermore, we find that both proteins are important for cohesin acetylation in G1. These results indicate that ESCO1 and CTCF function together to regulate cohesin's chromatin organisation activity. Our results underline that precise regulation of cohesin's residence time is key to how cells organise their genomes. They may also be of relevance for understanding the etiology of human cancers, in which STAG2 expression is often lost (Lawrence et al., 2014; Leiserson et al., 2015; Solomon et al., 2011).

Results

ESCO1 preferentially acetylates cohesin^{STAG1} during G1 phase

Acetylation of cohesin's SMC3 subunit during S phase is known to stabilise cohesive cohesin complexes on chromatin (Ladurner et al., 2016), but acetylated SMC3 (SMC3ac) can also be detected in quiescent (G0) cells and in cells in G1 phase, where no cohesive cohesin exists (Busslinger et al., 2017; Minamino et al., 2015; Whelan et al., 2012). To address possible functions of SMC3 acetylation in G1 phase we first analysed which cohesin complexes are modified during the cell cycle by ESCO1 and ESCO2. In this and subsequent experiments we detected acetylated SMC3 by using monoclonal antibodies that specifically recognise SMC3 which is acetylated either singly on K106 or doubly on K105 and K106 (Nishiyama et al., 2010). Immunoblot analyses of chromatin fractions isolated from synchronised HeLa cells confirmed that

SMC3 acetylation is detectable throughout the cell cycle, as is ESCO1, whereas ESCO2 expression is confined to S phase (*Figure S1A; lanes 2-5*). As predicted from these results, depletion of ESCO1 by RNA interference (RNAi) reduced SMC3 acetylation in G1, whereas depletion of ESCO2 had little effect in this cell cycle phase (*Figure S1B*; note that contrary to the situation in G1, in G2 both ESCO1 and ESCO2 contribute to SMC3 acetylation; see also Nishiyama et al., 2010). As reported by Minamino et al. (2015*),* cohesin acetylation in G1 phase is therefore predominantly mediated by ESCO1.

Unexpectedly, however, we found that cohesin^{STAG1} and cohesin^{STAG2} were acetylated to different extents in G1. When cohesin^{STAG1} and cohesin^{STAG2} were isolated from G1 chromatin fractions by immunoprecipitation using antibodies that recognise the endogenous proteins, more acetylated SMC3 was detected in the former than the latter sample (*Figure 1A; compare lanes 4 and 6*). These samples had been normalized to SMC3 to account for the three-fold higher abundance of cohesin^{STAG2} relative to cohesin^{STAG1} on chromatin (Holzmann et al., 2019) but similar results were also obtained when STAG1 and STAG2 immunoprecipitates from the same number of cells were analysed without SMC3 normalisation. Even though under these conditions more cohesin^{STAG2} was present than cohesin^{STAG1}, less acetylated SMC3 was detected in cohesin^{STAG2} than in cohesin^{STAG1} (*Figure S1C*). In contrast, cohesin^{STAG1} and cohesin^{STAG2} contained similar amounts of acetylated SMC3 in G2 phase (*Figure 1A; compare lanes 5 and 7*). These results indicate that SMC3 acetylation in G1 occurs preferentially on cohesin^{STAG1}.

Supporting this hypothesis, we observed that STAG1 depletion by RNAi reduced SMC3 acetylation in G1 to a similar extent as depletion of ESCO1, whereas STAG2 depletion had little effect (*Figure 1B*). This experiment also revealed that more STAG1 accumulates on chromatin in STAG2-depleted cells than in control-depleted cells (*Figure 1B*, lane 4). We also observed elevated levels of STAG1 in whole cell lysates prepared from STAG2-depleted cells (*Figure S1D, compare lanes 1 and 10*), implying that cells compensate for loss of STAG2 by increasing STAG1 levels by an unknown mechanism. However, this compensation is only partial as SCC1 levels were approximately three-fold lower in STAG2 depleted cells than in control cells (*Figure S1D*; compare lanes 1 and 10). This difference is important for the interpretation of Hi-C results which will be described below.

To determine the acetylation levels of cohesin^{STAG1} and cohesin^{STAG2}, we used label-free quantitative mass spectrometry (qMS). To be able to isolate cohesin^{STAG1} and cohesin^{STAG2} under comparable conditions we generated HeLa cells in which all STAG1 or STAG2 alleles were modified using CRISPR-Cas9-mediated genome editing to encode enhanced green fluorescent (EGFP) fusion proteins (for characterization of these cell lines, see *Figure S2A*) and isolated cohesin^{STAG1} and cohesin^{STAG2} using antibodies to GFP *(Figure 1C)*. This analysis indicated that in G1 cohesin^{STAG1} contains four times more acetylated SMC3 than cohesin^{STAG2} (*Figure 1D*).

A subpopulation of cohesin^{STAG1} associates stably with chromatin during G1 phase

Because SMC3 acetylation stabilizes cohesin on chromatin in S and G2 phase (Ladurner et al., 2016), we tested whether acetylated cohesin^{STAG1} also has a longer residence time on chromatin in G1 than the less acetylated cohesin^{STAG2} complexes. To analyse the chromatin binding dynamics of cohesin^{STAG1}, we performed inverse fluorescence recovery after photobleaching (iFRAP) in STAG2-depleted G1 cells that expressed a GFP-tagged version of SMC3 (SMC3-LAP; for depletion efficiency, see *Figure S2B*). Since STAG proteins are required for cohesin's association with chromatin (Roig et al., 2014), the behaviour of SMC3-LAP in these cells should predominantly reflect the behaviour of cohesin^{STAG1}. Conversely, to analyse cohesin^{STAG2} we analysed SMC3-LAP in cells depleted of STAG1. These experiments confirmed previous observations (Gerlich et al., 2006) that in control cells most cohesin interacts with chromatin dynamically

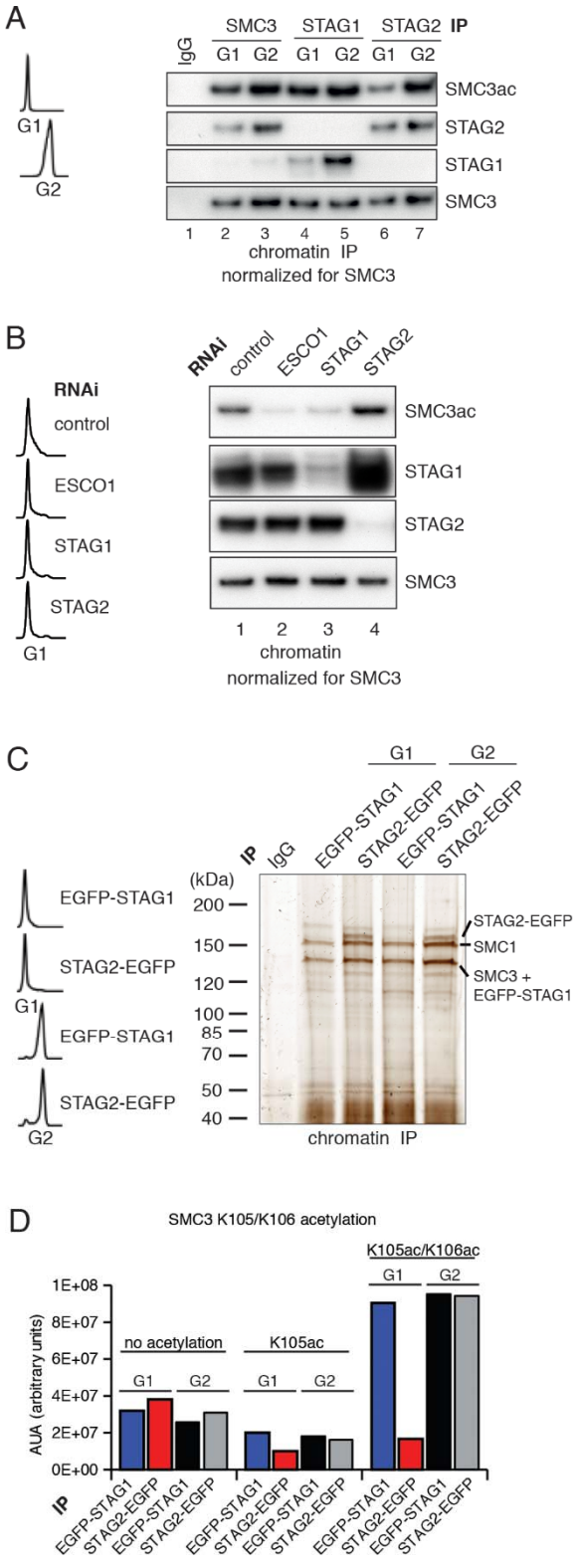


Figure 1. ESCO1 preferentially acetylates cohesin^{STAG1} during G1 phase

A. Immunoblot analysis of SMC3, STAG1 and STAG2 immunoprecipitates obtained from chromatin extracts of cells synchronized in G1 and G2. Immunoprecipitated material was normalized to SMC3 levels and immunoblotting was performed using the indicated antibodies. Flow cytometry profiles are shown on the left.

B. Flow cytometry and chromatin extract immunoblot analysis of cells synchronized in G1 and depleted of the proteins indicated. Chromatin extracts were normalized relative to SMC3. Note that SMC3ac levels are decreased after ESCO1 and STAG1 depletion and increased after STAG2 depletion.

C. SDS-PAGE analysis of EGFP-STAG1 and STAG2-EGFP immunoprecipitates obtained from chromatin extracts of cells synchronized in G1 and G2. Immunoprecipitations were performed using anti-GFP antibodies and were analysed with silver staining. Flow cytometry profiles are shown on the left.

D. Percentage of non-acetylated peptides (0 Ac), peptides acetylated at position K105 (K105) and peptides acetylated peptide at positions K105 and K106 (K105/106) was determined by quantitative mass spectrometry from material immunoprecipitated using anti-GFP antibodies from EGFP-STAG1 and STAG2-EGFP cells synchronized in G1 and G2. Peptide percentage was calculated relative to the total number of SMC3 peptides. Note that part of this experiment has been previously published in Ivanov at al., 2018; Figure S4.

with a residence of 13 min (*Figure 2A, B and F*). Similar results were obtained in cells depleted of STAG1, implying that in G1 most cohesin^{STAG2} interacts with chromatin dynamically (*Figure 2A, B and E*). In contrast, in cells depleted of STAG2, the SMC3-LAP signal equilibrated much more slowly between bleached and unbleached regions (*Figure 2A and B*) and the resulting iFRAP curve could only be fitted to a biexponential function (*Figure S2C*). This indicates that cohesin^{STAG1} complexes exist in two distinct populations which interact with chromatin differently. Analysis of these data revealed that most cohesin^{STAG1} (63%) is bound to chromatin dynamically with a residence time of 7 min, but that a smaller subpopulation of cohesin^{STAG1} (37%) is stably associated with chromatin with a residence time of 3 h (*Figure 2C and D*). The latter residence time is comparable to that observed for cohesin after depletion of WAPL (*Figure 2A-D;* Tedeschi et al., 2013), implying that the stably chromatin bound cohesin^{STAG1} complexes are protected from release by WAPL.

To analyse the stability of cohesin^{STAG1} and cohesin^{STAG2} on chromatin in G1 more directly, we performed iFRAP experiments in CRISPR-generated EGFP-STAG1 and STAG2-EGFP knock-in cell lines (*Figure 2G–L*, *Figure S2A and D*). As predicted from our STAG2 depletion experiments in SMC3-LAP cells, the EGFP-STAG1 iFRAP curve could only be fitted to a biexponential function. A subpopulation (33%) of EGFP-STAG1 was bound to chromatin stably (*Figure 2I*) with a residence time of 5 h (*Figure 2J*). The remaining 67% of EGFP-STAG1 bound dynamically to chromatin with a residence time of 15 min (*Figure 2K).* In contrast, the STAG2-EGFP iFRAP curve could be fitted to a single exponential function, with a dynamic residence time of five minutes (*Figure 2L*). A stably bound fraction of EGFP-STAG1 but not of STAG2-EGFP could also be detected in FRAP experiments, in which, compared to iFRAP, much smaller nuclear volumes are photobleached and therefore fluorescence recovery occurs on shorter time scales, resulting in higher temporal resolution (*Figure S3A-D*). Importantly, these differences between STAG1 and STAG2 were not caused by N-terminal tagging of STAG1 since we obtained similar results in an independent cell line, in which STAG1 was C-terminally tagged with EGFP (*Figure S3*; for technical reasons we were able to generate this cell line only late during this study, which is why earlier experiments were performed with a version of STAG1 which is tagged on its N-terminus, i.e. differently than the C-terminally tagged STAG2). These results indicate that a small subpopulation of cohesin^{STAG1} stably associates with chromatin in G1 in HeLa cells. We suspect that previous studies (Gerlich et al., 2006) failed to detect these complexes because they only represent about 9% of all cohesin complexes (37% of cohesin^{STAG1}; *Figure 2C*, which represents 25% of all cohesin; Holzmann et al., 2019).

Since this stable population of G1 cohesin has not previously been described, we tested whether such a population also exists in mammalian cells other than HeLa. We therefore performed Stag2 RNAi in immortalised mouse embryonic fibroblasts (iMEFs) and monitored the fluorescence recovery of Scc1-LAP (*Figure S4A–F*) or Smc1-LAP (*Figure S4G–L*) that were expressed from bacterial artificial chromosomes (BACs; *Table 1*). In contrast to our observations in HeLa cells, in iMEFs we were able to detect stably chromatin bound cohesin in G1 even without Stag2 depletion, perhaps because in these cells cohesin^{Stag1} represents 33 % of total cohesin (Remeseiro et al., 2012). We observed that approximately 20 % of Scc1-LAP and 12 % of Smc1-LAP were stably bound to chromatin in G1 in these cells (*Figure S4D and J*) with a residence time of 3 h (Scc1-LAP; *Figure S4F*) and 5 h (Smc1-LAP; *Figure S4L*). Following Stag2 depletion the stably bound fractions of cohesin increased to 35 % in both cell lines, indicating that also in MEFs predominantly cohesin^{Stag1} stably binds to chromatin (*Figure S4D and J*).

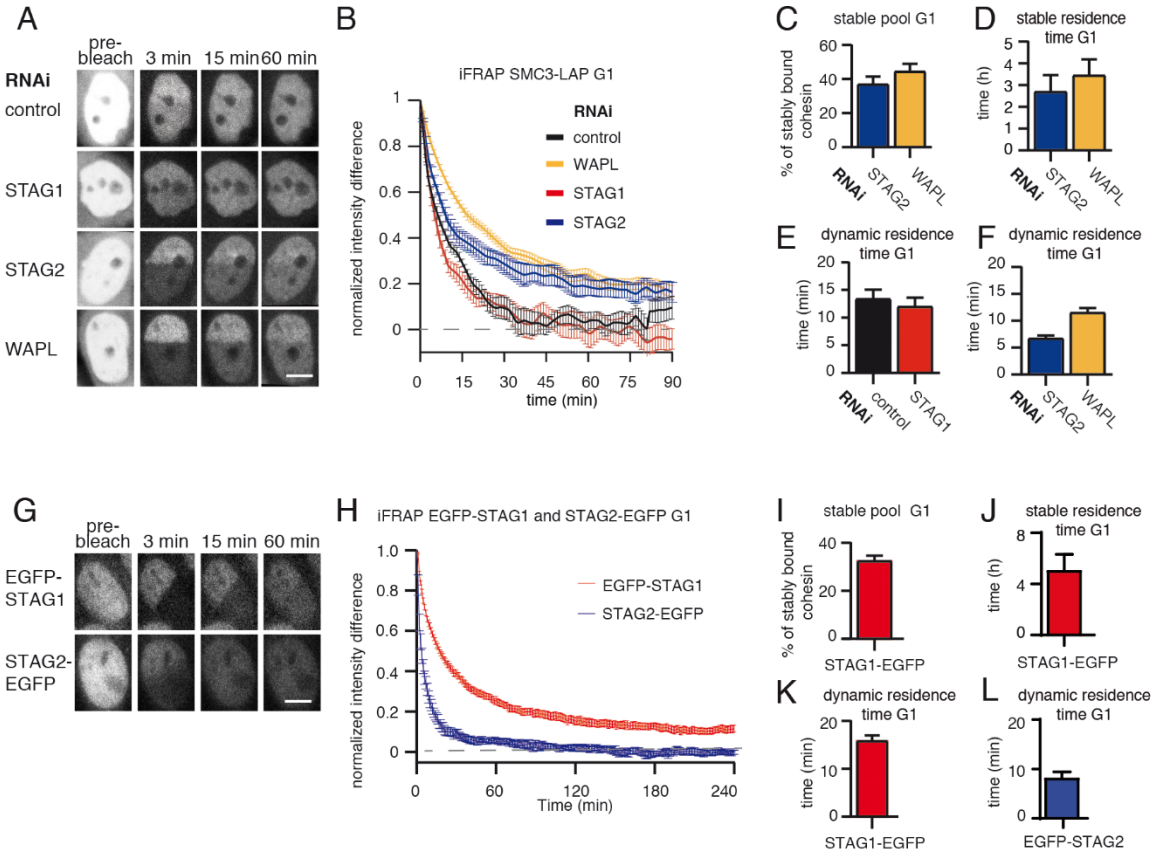


Figure 2. A subpopulation of cohesin^{STAG1} associates stably with chromatin during G1 phase

A. Images of inverse fluorescence recovery after photobleaching (iFRAP) experiments in SMC3-LAP cells synchronized in G1 and depleted of the indicated proteins by RNAi. Scale bar, 10 µm. Half of the nuclear SMC3-LAP fluorescent signal was photobleached and the mean fluorescence in the unbleached and bleached regions was monitored by time-lapse microscopy.
B. Graph depicting the mean normalized difference in fluorescence intensity between the bleached and unbleached regions from cells treated as described in A. Error bars denote standard error of the mean (s.e.m.), n >= 15 cells per condition.
C. Quantification of the fraction of nuclear SMC3-LAP that was stably chromatin-bound in cells synchronized in G1 and depleted of the indicated proteins by RNAi.
D. Quantification of the residence time of stably chromatin-bound SMC3-LAP in cells synchronized in G1 and depleted of the indicated proteins by RNAi.
E. Quantification of residence time of dynamically chromatin-bound SMC3-LAP in cells synchronized in G1 and depleted of STAG1 by RNAi. The numbers are derived from the single exponential fit.
F. Quantification of the residence time of dynamically chromatin-bound SMC3-LAP in cells synchronized in G1 and depleted of STAG2 and WAPL by RNAi. The numbers are derived from the bi-exponential fit.
G. iFRAP images in EGFP-STAG1 and STAG2-EGFP cells synchronized in G1. Scale bar, 10 µm.
H. Graph depicting the mean normalized difference in fluorescence intensity between the bleached and unbleached regions from cells treated as in G. Error bars denote standard error of the mean (s.e.m.), n = 10 cells per condition.
I. Quantification of the fractions of nuclear EGFP-STAG1 that was stably chromatin-bound in G1 cells.
J. Quantification of the residence time of stably chromatin-bound EGFP-STAG1 in G1 cells.
K. Quantification of the residence time of dynamically chromatin-bound EGFP-STAG1 in G1 cells. The numbers are derived from the biexponential fit.
L. Quantification of the residence time of dynamically chromatin-bound STAG2-EGFP in G1 cells. The numbers are derived from the single exponential fit.

The stable association of cohesin^{STAG1} with chromatin depends on ESCO1 and CTCF, as does SMC3 acetylation

Since stabilisation of cohesin in S and G2 depends on ESCO proteins and sororin (Ladurner et al., 2016; Schmitz et al., 2007), we tested whether the same proteins are also required for stabilisation of cohesin^{STAG1} in G1 (note that even though sororin levels are very low in G1, some sororin can be detected during this cell cycle phase; *Figure S2* in Nishiyama et al., 2010). To this end, we depleted sororin or ESCO1 by RNAi in G1 (*Figure 3A*) and measured the recovery of EGFP-STAG1 using iFRAP. Sororin depletion did not affect the chromatin binding dynamics of cohesin^{STAG1} (*Figure 3B and C*), as one might have expected given the low levels of sororin in G1 (Nishiyama et al., 2010; Rankin et al., 2005). This was not due to insufficient depletion of sororin because the same siRNA oligomer reduced the stable binding of EGFP-STAG1 to chromatin in G2 (*Figure S5F-J*). In contrast, depletion of ESCO1 converted most stably bound cohesin^{STAG1} complexes into dynamic ones (*Figure 3B and C*). This result was also observed using a second siRNA oligomer targeting ESCO1 (*Figure S5A–E*), suggesting that acetylation of cohesin^{STAG1} complexes is required for their stable association with chromatin.

Unexpectedly, while characterising our cohesin^{STAG1} and cohesin^{STAG2} immunoprecipitates, we noticed that CTCF was more abundant in the cohesin^{STAG1} sample, as detected by immunoblotting (*Figure 3D*) and label-free quantitative mass spectrometry *(Figure S6)*. We therefore tested whether CTCF is also required for stable binding of cohesin^{STAG1} to chromatin. To our surprise, CTCF depletion indeed reduced stable chromatin binding of EGFP-STAG1 to a degree similar to that observed following ESCO1 depletion (*Figure 3F and G*). A similar result was obtained by performing FRAP in a cell line in which STAG1 was tagged with EGFP and STAG2 was tagged with red fluorescence protein (RFP) at their endogenous loci (*Figure S7A-D*; for a characterisation of this cell line see *Figure S2A*). Also, in this cell line, STAG1 stability on chromatin was significantly reduced following CTCF depletion (*Figure S7A and B*). The dynamic residence time of STAG2 also decreased from 5 min to 3 min (*Figure S7A and C*), indicating that CTCF also prolongs the residence time of cohesin^{STAG2}, although to a much lesser extent than the residence time of cohesin^{STAG1}.

Interestingly, we also found that depletion of CTCF strongly reduced SMC3 acetylation levels (*Figure 3E*). as we had previously observed in primary Ctcf “knockout” MEFs arrested in G0 (Busslinger et al., 2017). This indicates that CTCF is also required for cohesin acetylation in G0 and G1.

Acetylated cohesin is enriched at loop anchors

Cohesin’s residence time on chromatin is thought to determine the lifetime and distance of chromatin interactions (Fudenberg et al., 2016; Gassler et al., 2017; Haarhuis et al., 2017; Wutz et al., 2017). Because some cohesin^{STAG1} complexes have long chromatin residence times, we analysed whether cohesin^{STAG1} complexes contribute to chromatin architecture differently during G1 than cohesin^{STAG2} complexes, which have short residence times.

For this purpose, we first analysed in ChIP-seq experiments where in the genome cohesin^{STAG1} is enriched compared to cohesin^{STAG2}. To avoid artefacts caused by the use of different antibodies, we used GFP antibodies in cell lines in which endogenous STAG1 and STAG2 had been tagged with EGFP (*Figure S2A*). This revealed that most STAG1 and STAG2 peaks overlapped (74%; *Figure 3H and I, left panel*). As recently reported for mouse cells (Kojic et al., 2018), sites at which predominantly STAG2 was found overlapped more frequently with transcription start sites (TSSs; 11.9%) and enhancers (5.5%) than STAG1-only sites, of which only 2% and 0.05% overlapped with TSSs and enhancers, respectively (*Figure S8A*). Kojic et al. *(2018)* also reported that STAG1-only sites overlap with CTCF sites more frequently than STAG2-only sites.

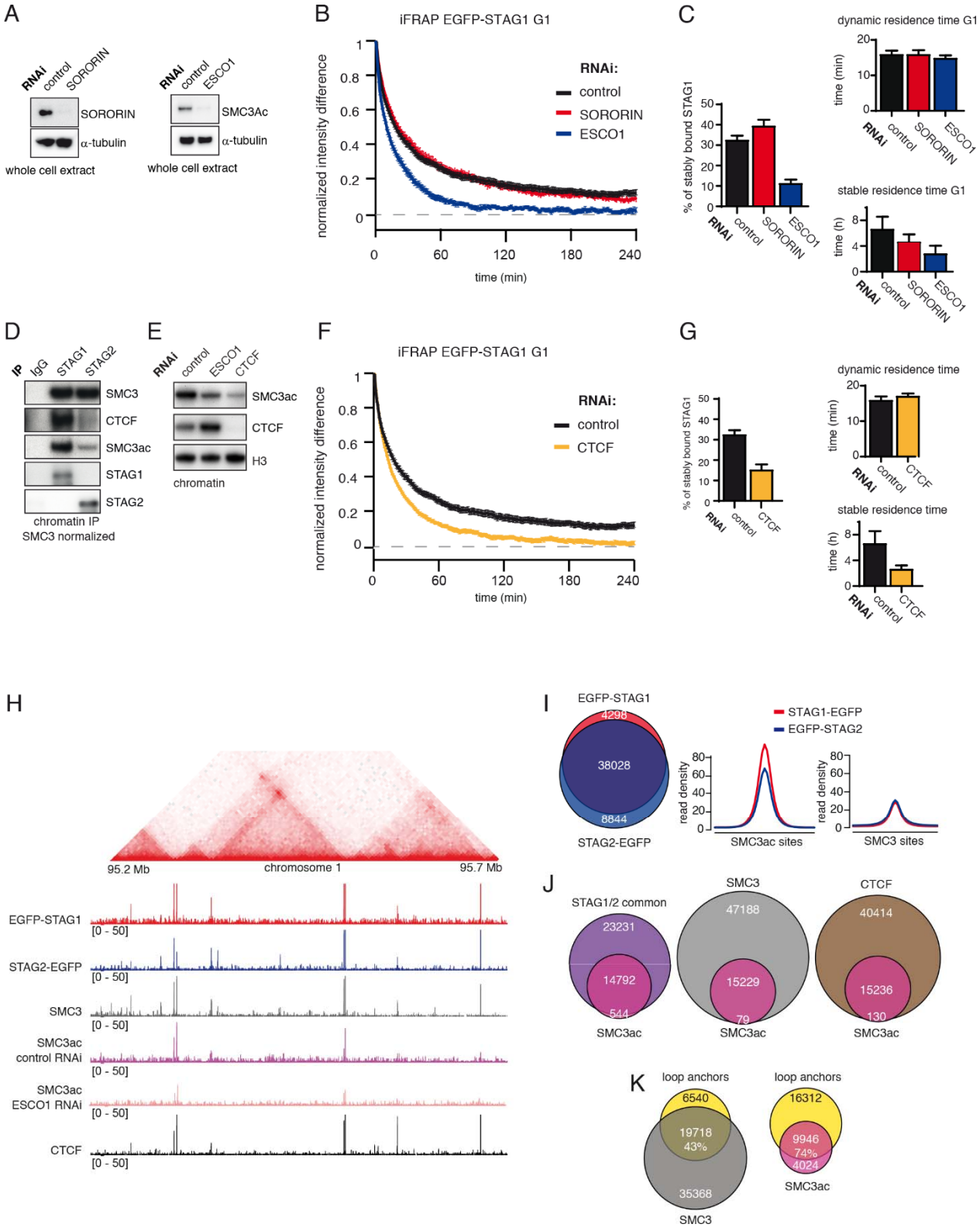


Figure 3. The long chromatin residence time of cohesin^{STAG1} depends on ESCO1 and CTCF, as does SMC3 acetylation

A. Immunoblot analysis of whole cell extract from cells depleted of sororin or ESCO1. α-tubulin was used as a loding control.
B. Graph depicting the mean normalized difference in EGFP-STAG1 fluorescence intensity between the unbleached and bleached regions following iFRAP in G1 cells and depleted of the indicated proteins by RNAi. Error bars denote standard error of the mean (s.e.m.), n = 10 cells per condition.
C. Quantification of the fraction of nuclear EGFP-STAG1 that was stably chromatin-bound in cells synchronized in G1 and depleted of the indicated proteins by RNAi. Quantification of dynamic and stable residence time of EGFP-STAG1 upon RNAi treatment for indicated proteins is shown on the right.
D.Immunoblot analysis of STAG1 and STAG2 chromatin immunoprecipitates obtained from cells synchronized in G1. Immunoprecipitations were performed using control-IgG, anti-STAG1 and anti-STAG2 antibodies (shown on the left). Immunoprecipitated material was normalized to SMC3 levels and immunoblotting was performed using the indicated antibodies.
E. Immunoblot analysis of chromatin extracts from control-, ESCO1- and CTCF-depleted HeLa cells synchronized in G1. Immunoblotting was performed using the indicated antibodies. Note that acetylation levels were decreased following depletion of CTCF. H3 was used as a loading control.
F. Graph depicting the mean normalized difference in EGFP-STAG1 fluorescence intensity between the bleached and unbleached regions following iFRAP in cells synchronized in G1 and depleted of the indicated proteins by RNAi. Error bars denote standard error of the mean (s.e.m.), n = 10 cells per condition.
G. Quantification of the fraction of nuclear EGFP-STAG1 that was stably chromatin-bound in cells synchronized in G1 and depleted of the indicated proteins by RNAi. Quantification of dynamic and stable residence time of EGFP-STAG1 in control and CTCF depleted cells by RNAi is shown on the right.
H. Coverage-corrected Hi-C contact matrix of the 95.2 – 95.7 Mb region of HeLa chromosome 1. ChIP-seq signals for EGFP-STAG1, STAG2-EGFP, SMC3, SMC3ac and CTCF are shown below the Hi-C contact matrix. STAG1 and STAG2 immunoprecipitation was performed using anti-GFP antibodies. Note that the SMC3ac signal is decreased following ESCO1 depletion, confirming the specificity of the anti-SMC3ac antibody.
I. Venn diagram illustrating the genome-wide co-localisation between STAG1 and STAG2 ChIP-seq signals. STAG1/2 immunoprecipitations from EGFP-STAG1 and STAG2-EGFP cells were performed using anti-GFP antibodies. Average read density plots for EGFP-STAG1 and STAG2-EGFP at SMC3ac and SMC3 sites is shown on the right.
J. Venn diagrams illustrating the genome-wide overlap between SMC3ac and STAG1/2 common sites, SMC3 and CTCF ChIP-seq sites respectively.
K. Venn diagrams illustrating the genome-wide overlap between the loop anchors determined using the Hi-C map shown in (H) and SMC3 (left panel) and SMC3Ac (right panel) ChIPseq binding sites.

Although we observed a similar tendency, we found that most STAG1-only and STAG2-only sites did not overlap with CTCF, whereas most common sites did *(Figure S8B)*.

Next, we determined where in the genome stably chromatin bound cohesin^{STAG1} complexes are enriched. To address this, we took advantage of our finding that many if not all of these complexes are acetylated on SMC3 *(Figure1).* We could detect acetyl-SMC3 at 24% of all SMC3 peaks (hereafter called pan-SMC3), at 39% of common STAG1 and STAG2 peaks and at 27% of CTCF peaks. Conversely, practically all SMC3-ac overlapped with SMC3, common STAG1 and STAG2 sites and CTCF (*Figure 3J*) but did not overlap with STAG1-only or STAG2-only sites *(Figure S8C)*. At the common STAG1 and STAG2 sites at which acetyl-SMC3 could be detected, the read density of STAG1 was higher than that of STAG2 (*Figure 3I, right panel*), consistent with a longer residence time and thus higher enrichment of some of the cohesin^{STAG1} complexes. Importantly, the number of acetyl-SMC3 peaks was reduced by ESCO1 depletion from 15,229 to 8,850, indicating that this antibody preferentially recognizes the acetylated form of SMC3 and not just unmodified SMC3 with reduced affinity (for a representative example, see *Figure 3H*).

We then used Hi-C to generate high resolution genome architecture maps in wild type HeLa cells (Hi-C map 1 in Table 2, 1.47 billion read pairs). By comparing the ChIP-seq profiles of pan-SMC3 and acetyl-SMC3 with these maps, we observed that acetyl-SMC3 was more frequently found at loop anchors (74%) than pan-SMC3 (43%; *Figure 3H and K*). This difference is almost certainly an underestimate of the specific enrichment of acetyl-SMC3 at loop anchors, as the pan-SMC3 profile presumably represents the sum of

unmodified and acetylated SMC3, and because due to the resolution of our Hi-C maps, the loop anchors were on average 9 kb long and contained on average three SMC3 peaks and one acetyl-SMC3 peak. In other words, an even higher resolution Hi-C map would be expected to reveal an even stronger enrichment of acetyl-SMC3 at loop anchors.

Cohesin^{STAG1} generates long chromatin interactions

Our data obtained so far indicated that, in G1, ESCO1 predominantly acetylates cohesin^{STAG1}, that some of these complexes bind to chromatin stably in an ESCO1 dependent manner, and that acetylated cohesin is enriched at loop anchors. We therefore hypothesized that cohesin^{STAG1} might be particularly important for forming long chromatin interactions.

To test this possibility, we depleted STAG1 or STAG2 by RNAi in cells synchronized in G1 (*Figure S9A*) and generated high-resolution Hi-C maps (Hi-C maps 2 and 3 in Table 2, >1.33 billion read pairs each). This revealed that following STAG1 depletion, i.e. under conditions in which predominantly cohesin^{STAG2} was present, the interactions within TADs and the number of loops decreased (*Figure 4A and Figure S9D*). In contrast, following STAG2 depletion, i.e. in the presence of cohesin^{STAG1}, new loops became detectable (*Figure 4A and Figure S9D*). This is remarkable given that our analysis of SCC1 had indicated that these cells contained three-fold less cohesin (see *Figure S1D* above). When we plotted the cumulative proportion of 7,177 loops only detected in STAG2-depleted cells as a function of their length, we found that they were longer than control-specific corner peak interactions of which we identified 14,726 (*Figure 4B and C*), confirming that cohesin^{STAG1} complexes form longer loops than cohesin^{STAG2}.

Corresponding changes could also be seen in contact probability plots *(Figure S9B).* Depletion of either STAG1 or STAG2 reduced the contact probability over short genomic distances (10–100 kb) but increased them over distances longer distances (>100 kb) compared to control cells, with STAG2 depletion having a much stronger effect. In control cells, the genomic distance with highest contact probability was around 200 kb, corresponding to the bulk of interactions that are associated with TADs. This peak shifted to ~800 kb following depletion of STAG2 (*FigureS9B and Table 2*).

Fluorescence *in situ* hybridization experiments support the population Hi-C data obtained from STAG1-depleted and STAG2-depleted cells

To test whether our observations made by Hi-C in populations of cells could be confirmed in individual cells, we performed fluorescence in situ hybridisation (FISH; *Figure S10*). We generated pairs of probes that hybridised to regions surrounding the bases of six loops with size of 0.8-1.5 Mb, as predicted by Hi-C (*Figure S10C*, left panels), and performed FISH with each probe pair in control-, STAG1- STAG2-, SCC1- and CTCF-depleted cells (*Figure S10A*; see *Figure S10B* for representative images from each experimental condition). We also performed the same analysis with a probe pair not predicted to span a loop (*Figure S10C, last example).* We used automated image analysis to measure the three-dimensional distance between each pair of probes in more than 100 cells in each experimental condition (number of cells analysed per condition and test of statistical significance between control and different conditions are listed in *Table 3*). This inter-probe distance was variable for each pair of probes, either reflecting technical variability, and/or indicating that the length of each predicted loop differed between cells, consistent with previous FISH and single-cell Hi-C studies (Flyamer et al., 2017; Nagano et al., 2013; Nora et al., 2017). In all six test loops, but not in the control genomic region, the inter-probe distance increased following depletion of SCC1 or

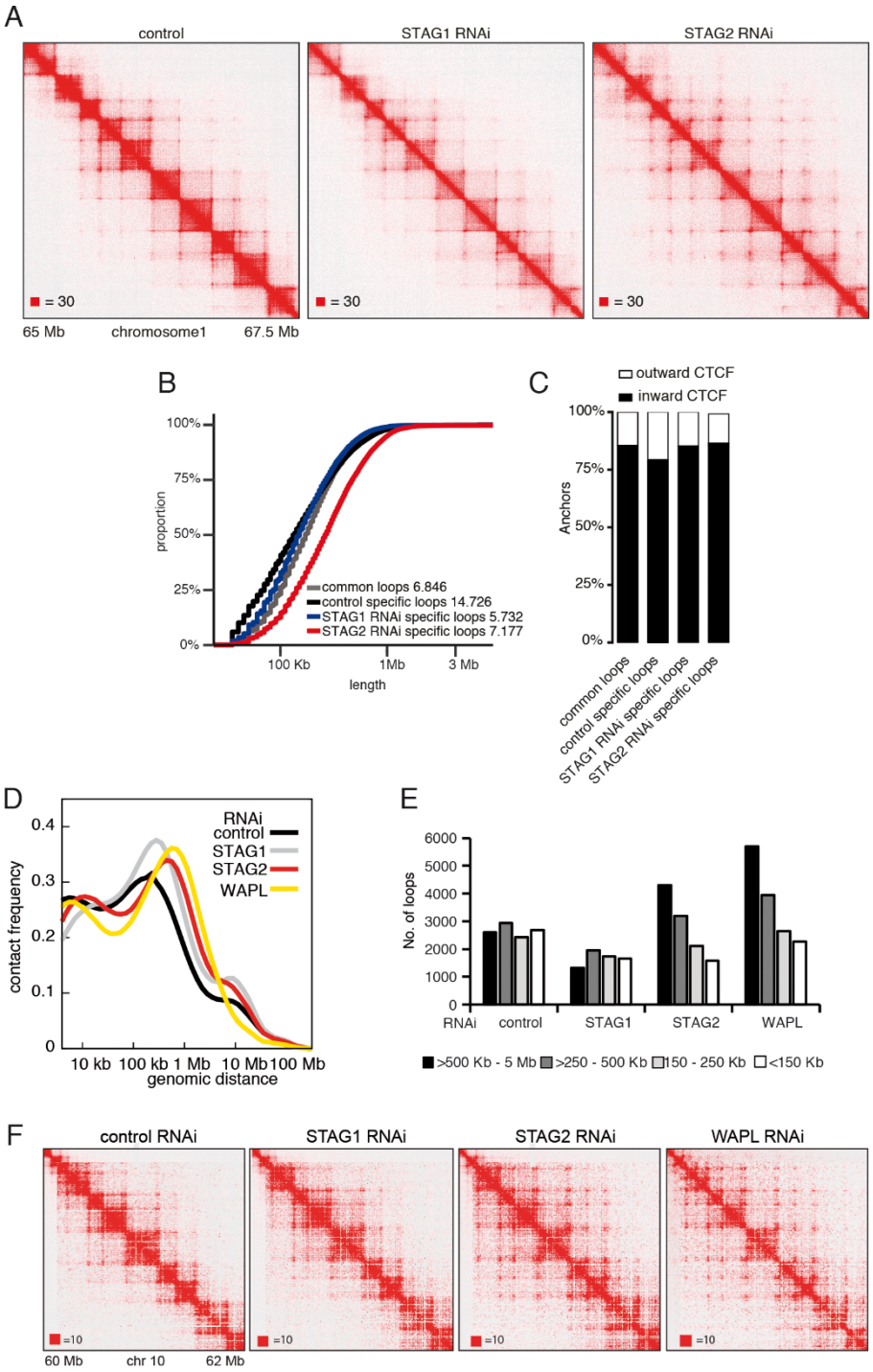


Figure 4. Cohesin^{STAG1} generates long chromatin interactions

A. Coverage-corrected Hi-C matrices (chr 1: 65 – 67.5 Mb) from control, STAG1 and STAG2-depleted cells.
B. Cumulative distribution of loop length determined by HiCCUPS in control, STAG1 and STAG2-depleted cells.
C. Quantification of the orientation of CTCF sites associated with Hi-C loops determined by HICCUPs.
D. Intra-chromosomal contact frequency distribution as a function of genomic distance using logarithmically increasing genomic distance bins, for control-, STAG1-, STAG2-, and WAPL-depleted HeLa cells.
E. Distribution of loop length in control-, STAG1-, STAG2- and WAPL-depleted HeLa cells. Loops identified by HiCCUPS.
F. Coverage-corrected Hi-C contact matrices of chromosome 10 (60 – 62.5 Mb), in control-, STAG1-, STAG2- and WAPL-depleted HeLa cells. Matrices were plotted using Juicebox.

following depletion of STAG1 in combination with STAG2. Depletion of CTCF also led to an increase in inter-probe distance in five out of six test loops, consistent with CTCF's proposed function as a boundary for loop formation. This indicates that this experimental setup can detect cohesin and CTCF-specific changes in chromatin architecture in single cells. Importantly, in five out of six test loops, depletion of STAG1 alone led to a greater increase in inter-probe distance than depletion of STAG2. This is consistent with our hypothesis that cohesin^{STAG1} is more important for generating longer-range chromatin loops than cohesin^{STAG2}.

Long-range chromatin interactions mediated by cohesin^{STAG1} are similar but not identical to those observed in WAPL depleted cells

Many of the loops that could only be detected in STAG2-depleted cells, i.e. were presumably mediated by cohesin^{STAG1}, had loop anchors in the outer boundaries of two or more adjacent TADs *(Figure 4A and Figure S9D)*, reminiscent of long-range chromatin interactions observed in WAPL depleted cells (Haarhuis et al., 2017; Wutz et al., 2017). We therefore compared Hi-C interactions in STAG2-depleted and WAPL-depleted cells. This revealed interesting similarities and differences.

First, we observed that most loops specifically detected in STAG2-depleted cells still "obeyed" the CTCF convergence rule (*Figure 4C*). This is in contrast to loops appearing after WAPL depletion, which frequently leads to violation of this rule (Haarhuis et al., 2017; Wutz et al., 2017). These results indicate that cohesin^{STAG1} forms long-range chromatin interactions that are anchored at convergent CTCF sites and raises the interesting possibility that WAPL is required for the CTCF convergence rule (see Discussion).

Second, to be able to compare the effects of STAG2 and WAPL depletion more directly, we generated libraries from control, STAG1-depleted and STAG2-depleted cells using the same experimental conditions and protocols as those previously used to study the consequences of WAPL depletion (*Figure 4D-F;* Hi-C maps 4,7,8 in Table 2, around 480 million read pairs each) and compared these to the previously generated Hi-C maps from WAPL-depleted cells (Wutz et al., 2017). Also in this case, the genomic distance with highest contact probability was around 200 kb in control cells, 300 kb in STAG1-depleted cells and 800 kb in STAG2-depleted cells (*Figure 4D*). Interestingly, the latter value is very similar to what had been observed in WAPL depleted cells, in which the genomic distance with highest contact probability was around 900 kb (*see Figure 5D in Wutz et al., 2017*). Likewise, the Hi-C maps obtained from STAG2-depleted cells resembled those depleted of WAPL, whereas Hi-C maps of STAG1-depleted cells were more similar to those of control cells (*Figure 4F).*

Third, aggregate peak analysis using loop coordinates identified in control or WAPL-depleted cells (Wutz et al., 2017) revealed that genome wide loops were weakest in STAG1-depleted cells, strongest in WAPL-depleted cells and of intermediate strength in STAG2-depleted cells (*Figure S11A*). Furthermore, loop calling by "Juicer tool" revealed that more loops >500 kb could be detected in STAG2-depleted cells than in control cells, whereas short loops (<250 kb) were less frequent. This phenotype is similar but not identical to the one observed in WAPL depleted cells, in which short loops (<250 kb) are similarly abundant as in control cells, whereas long loops (>250 kb) are more frequent. In contrast, interactions of all sizes were reduced in abundance in STAG1-depleted cells (*Figure 4E*).

Together, these observations support the notion that acetylated cohesin^{STAG1} complexes are protected from WAPL, therefore have longer residence times on chromatin and can form longer chromatin interactions, which, however, still obey the CTCF convergence rule.

In silico modelling indicates that cohesin^{STAG1}'s long chromatin residence time causes formation of long chromatin loops

To test the hypothesis that an increased chromatin residence time of cohesin^{STAG1} causes the formation of long chromatin loops we performed simulations *in silico*. For this purpose, we used a simplified hypothetical DNA sequence that contained three pairs of convergent CTCF sites and assumed three different extrusion times (*Figure 5A*). We also used molecular dynamics simulations to model the behaviour of chromatin in the presence of loop extrusion complexes and generated an *in silico* contact map of a region of human chromosome 9. This map resembled the Hi-C map of this region generated from control HeLa cells (*Figure 5B*, compare top and bottom panels). We next simulated the effect of altering the lifetime of loop extrusion complexes on DNA and found, consistent with results by Fudenberg et al. (2016) that longer lifetimes resulted in the generation of longer-range interactions, and *vice versa*. These *in silico* contact maps reproduced the changes we observed in cells following STAG1 and STAG2 depletion reasonably well. Thus, our experimental and simulation data are consistent with the hypothesis that cohesin^{STAG1} and cohesin^{STAG2} contribute to chromatin organisation differentially, and that, by virtue of its longer residence time on chromatin, cohesin^{STAG1} generates longer-range interactions than cohesin^{STAG2}.

ESCO1, like CTCF, regulates the loop formation activity of cohesin^{STAG1}

The hypothesis that stably chromatin bound cohesin^{STAG1} complexes form long chromatin loops predicts that also ESCO1 and CTCF are required for these interactions, since we had found that both of these are required for the long residence time of cohesin^{STAG1} (*Figure 3A-G*). To test this prediction, we co-depleted ESCO1 or CTCF together with STAG2, synchronized cells in G1 and performed Hi-C analysis (*Figure 6, Figure S12*; Hi-C maps 4,5,6,8,10 and 11 in Table 2, around 450 million read pairs each). Indeed, we found that the phenotypes caused by STAG2 depletion, namely the appearance of new loops and a decrease in signal intensity within TADs, were largely abolished by co-depletion of ESCO1 or CTCF (*Figure 6A and B*). This epistatic behaviour of ESCO1 and CTCF depletion over STAG2 depletion supports the hypothesis that the long residence time of cohesin^{STAG1} enables the formation of long chromatin loops in G1.

As controls, we also analysed cells from which only CTCF or ESCO1 had been depleted. Consistent with previous results obtained by auxin induced degradation (AID) of CTCF (Nora et al., 2017; Wutz et al., 2017), depletion of CTCF by RNAi resulted in a reduction in the number of detectable loops (*Figure 6A and B*), and a decrease in the TAD insulation score (*Figure 6C*) but did not abolish long-range chromatin interactions. To the contrary, contact probability analysis revealed that the interactions involved in TAD formation were longer in cells depleted of CTCF (*Figure S12B;* note that this particular effect was less pronounced after auxin induced CTCF degradation, perhaps because in these experiments CTCF levels were also reduced in control cells due to "leakiness" of the AID system; *Wutz et al., 2017*). As concluded previously (Nora et al., 2017; Wutz et al., 2017), these results suggest that CTCF is not required for long-range chromatin interactions *per se* but for specifying the loop anchors which mediate these, possibly by functioning as a boundary for loop extruding cohesin complexes. Remarkably, ESCO1 depletion had similar effects, i.e. a reduction in the number of detectable loops (*Figure 6A and B*), a decrease in the TAD insulation score (*Figure 6C*) and an increase in the length of chromatin interactions (*Figure S12B*). This raises the interesting possibility that ESCO1, like CTCF, is important for restricting cohesin's loop formation activity (see Discussion).

However, a comparison of Hi-C phenotypes between ESCO1-depleted and CTCF-depleted cells did not only reveal similarities but also differences. First, compared to both CTCF-depleted and control cells, contact frequencies were reduced in the 0–50 kb range following ESCO1 depletion (*Figure S12B*). Second, ESCO1

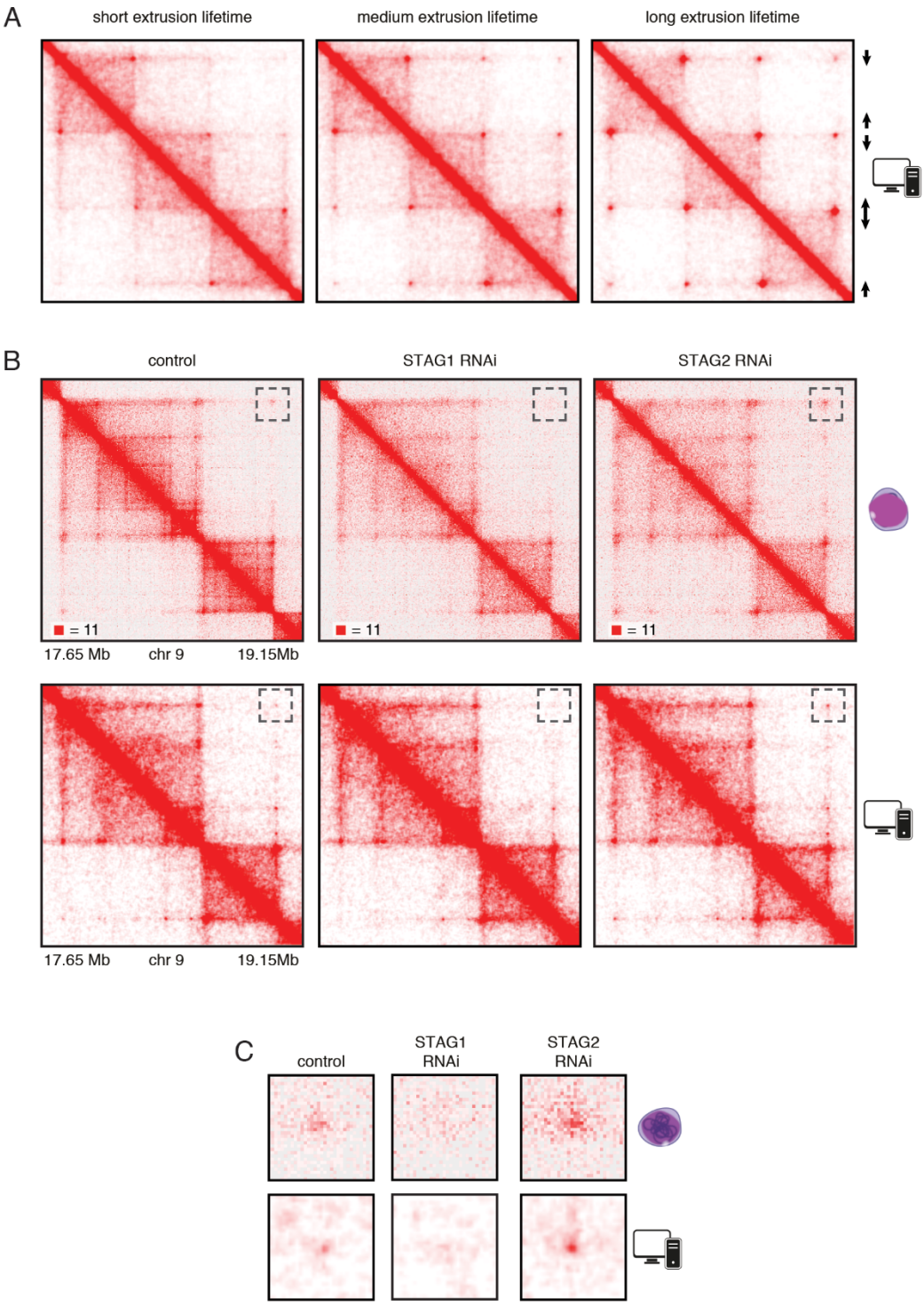


Figure 5. In silico modeling confirms that stably bound cohesin forms longer loops

A. Simplified simulation with three pairs of convergent CTCF sites to demonstrate the relationship between extrusion lifetime and loop size. Loop strength positively correlates with increased extrusion lifetime.
B. Coverage-corrected Hi-C matrices (chr 9: 17.65 – 19.15 Mb) from control, STAG1 and STAG2-depleted cells (upper panels). Hi-C simulations of the same genomic region were performed with short, medium and long extrusion complex lifetimes. Longer extrusion lifetimes generated Hi-C matrices that resembled those from STAG2-depleted cells.
C. Zoom in of the loops that are boxed in panel B.

depletion also resulted in a strong increase in contact probability around 20 Mb, the genomic distance associated with compartmentalisation (*Figure S12B*). Consistently, the analysis of whole chromosome Hi-C maps revealed that depletion of ESCO1, either alone or in combination with STAG2 depletion, enhanced the 'checkerboard' pattern indicative of compartmentalisation and led to a strong increase in interactions along the diagonal (around 10 Mb) (*Figure S12E*). Genome-wide aggregate analysis of 50 compartment categories ranging from strong B to strong A compartments confirmed this, showing increasing contact enrichment between similar compartment categories and a decreasing contact enrichment between dissimilar (e.g., strong A and strong B) compartment bins in both long *cis* (> 2 Mb) and *trans* interactions (*Figure S12F*). This phenotype is reminiscent of the increase in compartmentalisation observed following cohesin depletion (Flyamer et al., 2017; Rao et al., 2017; Schwarzer et al., 2017; Wutz et al., 2017), even though ESCO1-depleted cells contained as much cohesin on chromatin as the corresponding control of STAG2-depleted cells (*Figure S12A*, compare SCC1 signals in lanes 3 and 6, and in lanes 9 and 12). These observations raise the interesting possibilities that the ability of cohesin to suppress compartmentalisation depends on ESCO1, or that this acetyltransferase has additional functions in chromatin organisation that are independent of cohesin.

Discussion

Acetylated cohesin^{STAG1} complexes are protected from WAPL by CTCF and form long chromatin loops

In interphase cells, cohesin folds genomic DNA into thousands of loops which are thought to have both structural and regulatory functions. Little is known about the lifetime of these loops and how their formation and maintenance is controlled. It has generally been assumed that loops are short-lived dynamic structures because the cohesin complexes that form them interact with DNA only briefly, in mammalian cells on average for 8-25 minutes during G0 and G1 phase (Gerlich et al., 2006; Tedeschi et al., 2013), after which they are released by WAPL (Kueng et al., 2006). Despite this, long-range chromatin interactions as they can be detected by Hi-C change little over time (Nagano et al., 2017; Wutz et al., 2017), and recent evidence implies that some loops can persist for hours (Vian et al., 2018). It is not known whether these stable structures are maintained by dynamically exchanging cohesin complexes or by an unknown mechanism that would protect cohesin from release by WAPL. Precedence for the latter scenario comes from the observation that cohesin complexes that mediate cohesion in proliferating cells are protected from WAPL by acetylation of their SMC3 subunit (Ben-Shahar et al., 2008; Ünal et al., 2008), by subsequent recruitment of sororin and at mitotic centromeres also by shugoshin (Hara et al., 2014; Nishiyama et al., 2010). However, these cohesive complexes do not seem to participate in loop formation, as their stabilisation on chromatin does not detectably alter chromatin structure in G2 phase (Wutz et al., 2017), whereas experimental stabilisation of all cohesin complexes on chromatin does (Tedeschi et al., 2013; discussed in Holzmann et al., 2019). It has therefore remained unknown whether loop forming cohesin complexes can be protected from WAPL to extend the lifetime of chromatin loops.

Here we provide evidence that such a regulatory mechanism exists in human cells, since our FRAP and Hi-C experiments have identified a small subpopulation of cohesin^{STAG1} complexes that persist on chromatin for hours and contribute to the formation of long chromatin loops. Our results indicate that the stabilisation of these cohesin^{STAG1} complexes on chromatin depends on SMC3 acetylation, as does the stabilisation of cohesive cohesin in S and G2 (Ladurner et al., 2016). But in contrast to cohesive cohesin, loop forming cohesin^{STAG1} complexes can persist on chromatin for hours in the absence of sororin, as one might have

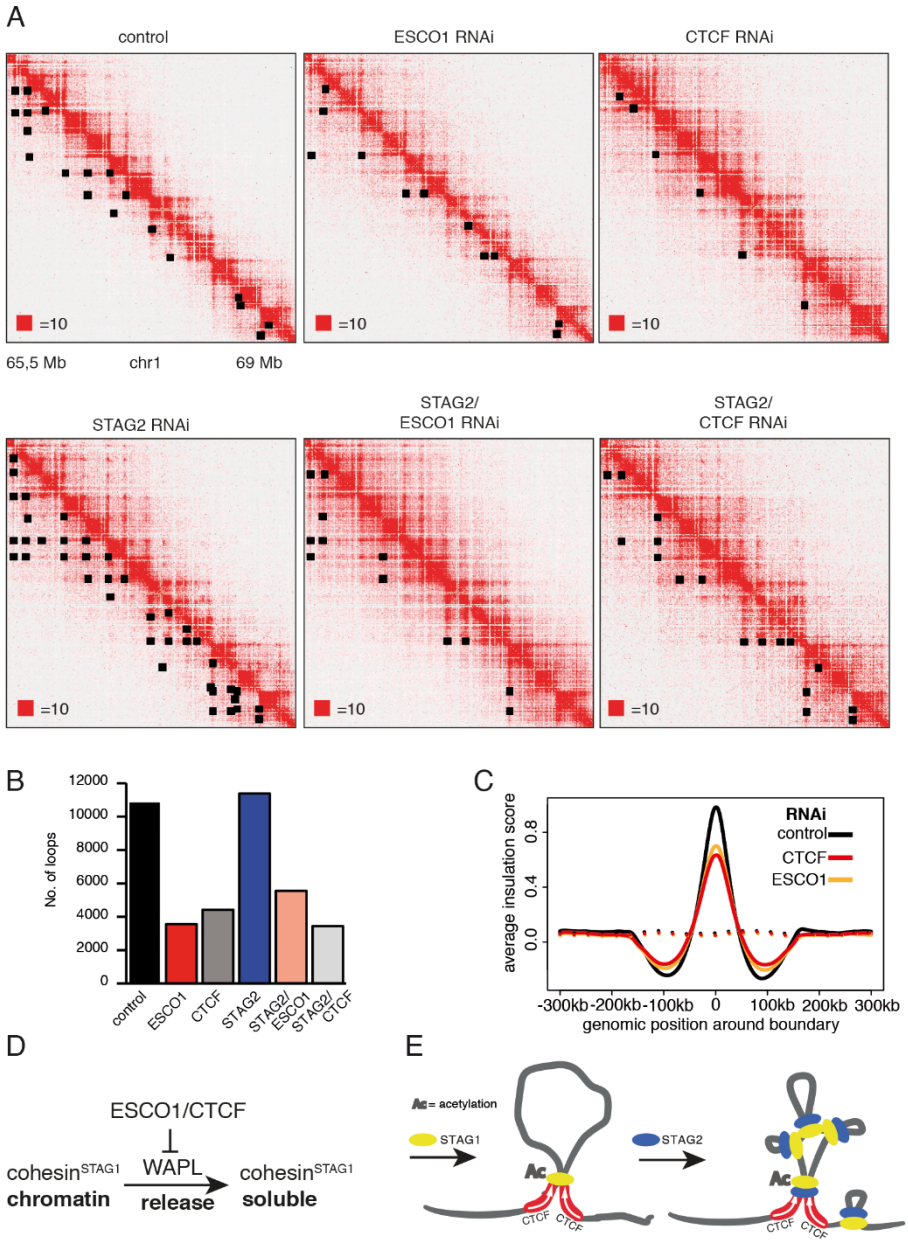


Figure 6. The function of long chromatin loops by cohesin^{STAG1} depends on CTCF and ESCO1

A. Coverage-corrected Hi-C contact matrices of chromosome 1 (65,5 – 69 Mb), in control-, ESCO1-, CTCF-, STAG2-, STAG2/ESCO1, and STAG2/CTCF-depleted HeLa cells. Loops identified in this region by HICCUPs are marked by black rectangles. Matrices were plotted using Juicebox.
B. Number of loops identified by HICCUPs in the matrices described in (A).
C. Average insulation score around TAD boundaries in control, ESCO1 and CTCF-depleted cells synchronised in G1. Dashed lines indicate the average insulation score between random positions.
D. Schematic model: ESCO1 and CTCF stabilize cohesin^{STAG1} by inhibiting WAPL.
E. Schematic model of nested loop extrusion. Cohesin^{STAG1} makes longer loops; cohesin^{STAG2} makes loops in and around cohesin^{STAG1} loops.

predicted since sororin is only present in proliferating cells, and in these almost exclusively from S phase until mitosis (Nishiyama et al., 2010; Rankin et al., 2005). Likewise, we suspect that shugoshin is dispensable for the long chromatin residence time of acetylated cohesin^{STAG1}, since shugoshin specifically

protects cohesive cohesin from WAPL at centromeres in mitosis (Hara et al., 2014). Unexpectedly, however, we found that CTCF is essential for the long residence time of cohesin^{STAG1} on chromatin, indicating that CTCF, like sororin and shugoshin, is a WAPL antagonist that can prevent cohesin from being released from DNA (*Figure 6D*).

The notion that cohesin^{STAG1} complexes are protected from WAPL by SMC3 acetylation and CTCF is supported by the observation that the chromatin residence time of these complexes is similarly long as the residence time of cohesin in WAPL depleted cells, and by our finding that cohesin^{STAG1} complexes form similarly long chromatin loops as cohesin does in the absence of WAPL. Interestingly, however, the extended loops formed by cohesin^{STAG1} differ in one important aspect from the loops that are formed by cohesin in the absence of WAPL, in that the former are typically anchored at convergent CTCF sites (this study) whereas the latter are often not (Haarhuis et al., 2017; Wutz et al., 2017). This difference implies that interactions between loop forming cohesin complexes and CTCF may not be sufficient for the CTCF convergence rule, and that instead WAPL may have a direct role in ensuring that loops are only anchored at convergent CTCF sites.

Our photobleaching experiments also revealed that most if not all cohesin^{STAG2} complexes, which are three fold more abundant in HeLa cells than cohesin^{STAG1} complexes (Holzmann et al., 2019), have short chromatin residence times in the range of minutes, implying that they are dynamically released from chromatin by WAPL. Accordingly, CTCF depletion only had a small effect on their chromatin residence time. However, this does not exclude the possibility that CTCF would also be able to protect cohesin^{STAG2} from WAPL under conditions where these complexes become stabilised on chromatin.

The existence of two different forms of cohesin with different chromatin residence times is reminiscent of the situation in mitotic chromosomes. Their structural organisation has been shown to depend on the sequential action of first condensin II to form large loops and then condensin I to form smaller, nested loops (Gibcus et al., 2018). It is unknown how interphase chromatin architecture is established upon exit from mitosis, but recent fluorescence correlation spectroscopy experiments indicate that STAG1 is recruited to chromatin earlier than STAG2 (Cai et al., 2018). Given our finding that cohesin^{STAG1} generates longer loops than cohesin^{STAG2}, it is tempting to speculate that cohesin^{STAG1} and cohesin^{STAG2} function analogously to condensin II and condensin I in mitosis, respectively, by forming loops within loops on interphase chromatin (*Figure 6E*).

Consistent with our findings, it has recently been reported that cohesin^{STAG1} and cohesin^{STAG2} contribute to chromatin differently (Casa et al., 2019; Viny et al., 2019; Cuadrado et al., 2019; Kojic et al., 2018), with cohesin^{STAG1} contributing more to TAD organisation than cohesin^{STAG2}. Cohesin^{STAG1} has also been shown to be more resistant to biochemical salt extraction from chromatin than cohesin^{STAG2} (Kojic et al., 2018), but whether this property is related to the long chromatin residence time of acetylated cohesin^{STAG1} reported here is unknown.

Both our study and that of Kojic et al., 2018 provide evidence that STAG1 and STAG2 are enriched at CTCF sites and enhancers, respectively. This raises the possibility that cohesin^{STAG2} might regulate promoter–enhancer interactions. STAG2 is one of only twelve genes known to be mutated in more than four major human cancer types (Lawrence et al., 2014). It is therefore possible that changes in gene expression following STAG2 mutation might be a common early event in human carcinogenesis.

ESCO1 is needed to constrict cohesin at CTCF sites

The acetylation that protects cohesive cohesin from WAPL from S phase until mitosis is thought to depend on both ESCO1 and ESCO2 (Ladurner et al., 2016; Nishiyama et al., 2010). In contrast, acetylation of cohesin during G1 is only mediated by ESCO1, since ESCO2 is absent during this phase of the cell cycle (Minamino et al., 2015; Whelan et al., 2012; *Figure S1A*). Our finding that ESCO1 is needed for a long chromatin residence time of a subpopulation of cohesin complexes and for their ability to form extended chromatin loops reveals for the first time a function of SMC3 acetylation in G0 and G1, and at the same time indicates that ESCO1 has a certain degree of substrate specificity, at least during G1 phase, by preferentially acetylating cohesin^{STAG1}.

Interestingly, our results revealed that ESCO1 might also have another function in chromatin organisation. ESCO1 depletion reduced the Hi-C dots and corner peaks specifically detected after STAG2 depletion, i.e. long chromatin loops formed by cohesin^{STAG1}, as one would have predicted given that ESCO1 is required for the long residence time of a subpopulation of cohesin^{STAG1}. However, in addition, ESCO1 also decreased TAD insulation, to an extent as seen after CTCF depletion. This implies that ESCO1 might have two functions in chromatin organisation: protection of cohesin^{STAG1} from WAPL, and a boundary function in loop extrusion. If ESCO1 has such a boundary function, this could help to explain why co-depletion of ESCO1 and ESCO2 with WAPL resulted in the formation of vermicelli in chicken DT40 cells, whereas in these cells WAPL depletion alone did not cause this phenotype (Kawasumi et al., 2017). It is possible that in these experiments WAPL depletion increased the residence time of cohesin on chromatin, whereas depletion of ESCO1 compromised the function of CTCF boundaries, so that extruding cohesin complexes could form longer loops and ultimately accumulate in vermicelli domains.

The hypothesis that ESCO1 contributes to boundary function could also explain why PDS5A and PDS5B are required for boundary function (Wutz et al., 2017) because SMC3 acetylation by ESCO1 and ESCO2 has been shown to depend on PDS5A and PDS5B (Minamino et al., 2015; for a similar dependency in yeast see Chan et al., 2013; Vaur et al., 2012). It is therefore possible that one function of PDS5 proteins in chromatin architecture is to facilitate SMC3 acetylation by ESCO1. In addition, PDS5 proteins might influence chromatin architecture by other mechanisms as they are also required for WAPL-mediated release of cohesin from chromatin (Chan et al., 2012; Ouyang et al., 2016; Shintomi and Hirano, 2009; Wutz et al., 2017). Interestingly, we also observed a reduction in SMC3 acetylation following depletion of CTCF, indicating that SMC3 acetylation also depends on CTCF, and that the functions of ESCO1 and CTCF at chromatin boundaries might be interdependent (Busslinger et al., 2017*;* this study).

Consistent with our results, a role for the yeast ortholog of ESCO1 and ESCO2 (Eco1) in constraining cohesin's ability to form mid-range chromatin interactions, together with a dual role for the PDS5A and PDS5B ortholog Pds5 in regulating Eco1 and Wapl was reported during preparation of this manuscript (Dauban et al., 2019). The molecular mechanisms through which STAG1 and STAG2 alter the properties of cohesin remain to be understood, but our results suggest that cohesin^{STAG1} is more frequently stabilised at CTCF sites via PDS5A, PDS5B and ESCO1 than cohesin^{STAG2}. The regulation of chromatin boundaries by cohesin acetyltransferases and PDS5 proteins may therefore be an evolutionarily conserved mechanism, which might be spatially controlled in mammalian cells by CTCF.

Materials and Methods

Cell culture, cell synchronization and RNA interference

HeLa Kyoto cells (Landry et al., 2013) were cultured in DMEM supplemented with 10% FBS, 0.2 mM glutamine and penicillin/streptomycin (Gibco). Cells were synchronized at early S phase by two consecutive rounds of treatment with 2 mM thymidine (Sigma) and released into fresh media for 6 h (G2) or 15 h (G1). Synchronization was assessed by flow cytometry after methanol fixation and propidium iodide staining as described (Ladurner et al., 2014). Cells were treated with 30 nM siRNAs as indicated using RNAiMax (Invitrogen) at 48 or 72 hours before downstream analyses. Pre-annealed 21 nucleotide RNA with 3' double thymidine overhangs (Elbashir et al., 2001) was purchased from Ambion. Sense sequences for control, CTCF ("#1") (Wendt et al., 2008), ESCO1, ESCO2 (Nishiyama et al., 2010), Wapl ("Wapl1"), STAG1 ("SA1"), STAG2 ("SA2") (Kueng et al., 2006), and sororin (Schmitz et al., 2007) were denoted previously. ESCO1 second siRNA was **a** pool of 4 siRNAs: GGAAAGAGCAAACGAGGUA, GGACAGAAUAGCACGUAAA, CUAGAAGAGACGAAACGAA, GGACAAAGCUACAUGAUAG.

Generation of cell lines

STAG1-EGFP, EGFP-STAG1, STAG2-EGFP, EGFP-STAG1-RFP-STAG2

Homology arms (0.6-1.5kb per arm) surrounding the start or stop codons of STAG1 and STAG2 were amplified from genomic DNA of HeLa Kyoto cells using primers identified by primer-blast (Ye et al., 2012) and cloned into vector pJet1.2 (Thermo Scientific K1232). EGFP or FLAG-mRFP coding sequences were introduced before the stop or after the start codon to generate homology-directed recombination (HDR) donor plasmids. CRISPR guide RNAs introducing nicks on either strand when bound to SpCas9(D10A) (Ran et al., 2013) were identified using crispr.mit.edu and cloned into plasmid pX335 (Addgene 42335). The following *genomic* sequences were targeted: ACAATACTTACTGTAACACtgg and TATTTTTTAAGGAAAATTTtgg (STAG1 N-terminus); TGAAGAAAATTTACAAATCtgg and TCTTCAGACTTCAGAACATagg (STAG1 C-terminus); ATTTACGTGGGTAAAATGGtgg and GAATATATTTCTGACATTGagg (STAG2 N-terminus); CACAGATTTAATTGTGTACtgg and CAGTACACAATTAAATCTGtgg (STAG2 C-terminus). HeLa Kyoto cells were transfected with two guide RNA and one HDR donor plasmid (or 4 guide RNA and two HDR donor plasmids for the EGFP-STAG1 FLAG-mRFP-STAG2 double tagged cell line) using Lipofectamine 2000 (Invitrogen 11668019). Cells were grown for 7-10 days before sorting single cells into 96 well plates. Homozygous targeting of genomic alleles was assessed by PCR and by immunoblotting after fractionation (Ladurner et al., 2016).

Scc1-LAP, Smc1-LAP

For generation of Scc1-LAP and Smc1-LAP immortalized mouse embryonic fibroblasts (iMEFs), primary mouse embryonic fibroblasts (pMEFs) were isolated from E13.5 embryos as described previously (E. Michalska, 2007). Immortalized mouse embryonic fibroblasts (iMEFs) were then generated by the 3T3 protocol. The LAP tag was introduced as described (Poser et al., 2008). Briefly, Smc3-LAP or Scc1-LAP BAC constructs were introduced using Fugene HD transfection reagents. Cells were then selected based on geneticin (G418) resistance and thereafter FACS sorted based on GFP expression levels.

FRAP

For FRAP of SMC3-LAP, cells expressing fluorescent cohesin subunits were grown on chambered coverglass (Nunc 155409) for 1-3 days while treated with siRNAs and thymidine as indicated. Cells were imaged at 37°C on a Zeiss LSM5 duo confocal microscope with 63x Plan-Apochromat objective and a 488

nm 100 mW diode laser for bleaching, or on a Zeiss LSM780 confocal microscope with 63x objective and bleaching with argon and diode lasers at 488 and 561 nm for dual color FRAP using CO2-independent media, or on an LSM880 confocal microscope (Carl Zeiss) with a 40 ×1.4 NA oil DIC Plan- Apochromat objective (Zeiss) in cell culture medium without riboflavin and phenol red at 5% CO2. Cells were either cell cycle synchronized as described above, or G1 and G2 phase were identified by nuclear and cytoplasmic distribution of DHB-mKate2 signals, respectively. Cycloheximide (1 ug/ml) was added before imaging to inhibit protein synthesis and contribution of new GFP expression to signal recovery.

For spot FRAP, a circular region (r = 2 um) was bleached 3 times. Recovery of fluorescence was recorded over 10 minutes and 300 frames at 2 second intervals and normalized to 10 pre-bleach frames and background and cellular fluorescence measured with Fiji (Schindelin et al., 2012). Recovery curves were analyzed using Berkeley Madonna (www.berkeleymadonna.com). Curves were fitted by an exponential function with variables for free and transiently chromatin associated (Ladurner et al., 2014), dynamically and stably chromatin bound cohesin (Gerlich et al., 2006). Relative fractions and their residence times (reciprocal of the dissociation constant) were averaged and plotted using Prism software (GraphPad).

Inverse FRAP was used to specifically measure dynamics of cohesin bound to chromatin over several hours. To this end for SMC3-LAP (which is present in the nucleus and cytoplasm), an area covering the cell body was bleached except for a semicircle corresponding to approximately half of the nucleus, and cells were imaged intermittently using a motorized stage. Recovery was recorded over 2-4 hours at 3-minute intervals and normalized as above. Curve fitting with single and bi-exponential function was used to deduce relative fractions and residence times of dynamic and stable cohesin on chromatin and plotted as above (see https://github.com/rladurner/STAG1/blob/master/curvefit.ipynb). For STAG1-EGFP and EGFP-STAG2 (which show only nuclear GFP signal), iFRAP photobleaching was performed in half of nuclear regions with 2 iterations of 488 nm laser at max intensity after acquisition of two images. Fluorescence was measured in bleached- and unbleached regions followed by background subtraction with 1 min interval. iFRAP curves were normalized to the mean of the pre-bleach fluorescent intensity and to the first image after photobleaching. Curve fitting was performed with single 18 exponential functions f(t) = EXP(-kOff1*t) or double exponential functions f(t) = a*EXP(-kOff1*t)+(1-a)*EXP(-kOff2*t) in R using the minpack.lm package (version 1.2.1). Dynamic and stable residence times were calculated from 1/kOff1 and 1/kOff2 respectively. Double exponential curve fitting was performed under constraint that 1/kOff1 and 1/kOff2 are in range between 1 min-40 min and 1.5 hr-15 hr respectively. Soluble fractions were estimated by the reduction of fluorescence signals in unbleached area after photobleaching.

Chromatin fractionation

Cells were extracted in a buffer consisting of 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 5 mM MgCl, 2 mM NaF, 10% glycerol, 0.2% NP40, 20 mM β-glycerophosphate, 0.5 mM DTT and protease inhibitor cocktail (Complete EDTA-free, Roche). Chromatin pellets and supernatant were separated and collected by centrifugation at 2,000 *g* for 5 min. The chromatin pellets were washed three times with the same buffer.

Antibodies

For western blotting the following rabbit antibodies were used: STAG1 (Peters laboratory, A823), STAG2 (Peters laboratory, A824), ESCO1 (Peters laboratory 782M), CTCF (Peters laboratory A992), sororin (Peters laboratory ID 953). The following commercial antibodies were used: SMC1 (Bethyl Laboratories, A300-055A), SMC3 (Bethyl Laboratories, A300-060A), SCC1 (EMD Millipore Corporation, 53A303), PDS5A (Bethyl Laboratories, A300-089A), α-tubulin (Sigma-Aldrich, T5168), histone H3 (Santa Cruz Biotechnology, sc-8654), phospho-histone H3 (Ser10) (Cell Signaling, #9701), PCNA (Santa Cruz, PC10), GFP (Roche,

clones 7.1 and 13.1; Abcam, ab13970). The following secondary antibodies were used for WB anti-rabbit or mouse Ig, HRP-linked whole antibody (GE Healthcare) and polyclonal rabbit anti-goat (Dako), mouse anti-acetyl SMC3 was a gift from K. Shirahige and guinea pig anti-ESCO2 was a gift from J. de Winter.

For ChIP-seq the following antibodies were used: GFP (Abcam, Ab290), STAG1 (Peters laboratory, A823), STAG2 (Peters laboratory, A824), SMC3 (Peters laboratory A941), CTCF (Merck Millipore,07-729)

Mass spectrometry

To generate a peptide spanning lysin 105 and 106 the immunoprecipitates were digested in solution with 400 ng Glu-C (Sequencing Grade, Roche) at 37C for 16 hours.
The mass traces for the peptide VSLRRVIGAKKD in its unmodified, singly acetylated and doubly acetylated form were extracted from the raw files using the program Qualbrowser which is part of the Xcalibur software (Thermo Scientific). To account for different amounts of SMC3 protein between the immunoprecipitates the peptide area values were normalized based on the sum of the 3 most intense unmodified SMC3 peptides.

DNA FISH

DNA probes (BACs, Fosmids and Cosmids) were ordered at BAC Resources PAC and purified by midi-prep purification kit (Quiagen). Probes were labelled by nick translation using 1-2 µg DNA per 50 µL reaction (Sigma Aldrich). Probes were fluorescently labelled using Alexa dyes (Alexa-488, Alexa-568). Every 5 µL of nick-translated probe was ethanol precipitated together with 1 µL of salmon sperm DNA, 3µL human Cot-1, 0.5 µL 3M sodium acetate and 60µL ethanol 100%. Probes were then snap frozen in liquid nitrogen and centrifuged at 13000 rpm for 10 min at 4 degrees. The supernatant was carefully removed and replaced with 200 µL of ethanol 70%, and the probe mixture was centrifuged at 13000 rpm for 10 min at 4 degrees. The supernatant was carefully removed and the pellet air dried protected from light. The pellet was resuspended in 5µL hybridization buffer (2xSSC, 20% w/v dextran sulfate, 50% formamide pH7) at 37 degrees for 10 minutes then denatured for 7 min at 80 degrees and incubated at 37 degrees for 30 min before use. HeLa cells were cultured, siRNA treated and synchronized on coverslips. Cells were quickly rinsed with PBS three times and fixed in a solution of PBS with 4% paraformaldehyde for 10 minutes at room temperature. Cells were then washed twice in PBS for 5 min each. Permeabilization of cells was performed in freshly made PBS, 0.5 Triton X-100 for 7 min on ice. Cells were washed twice with a solution of ethanol 70% for 5 min and dehydrated in 80%, 95% and 100% ethanol for 3 min each. Cells were then air dried and denatured in 50% formamide, 2xSSC adjusted at pH 7.2 for 30 min at 80 degrees. Cells were washed three times in cold 2xSSC. The coverslip was placed cell-side down onto the prepared fluorescently labelled probes on a slide and sealed with glue. Hybridization was performed overnight at 42 degrees in a dark and humid chamber. The glue was then removed carefully, and coverslips were placed cell-side up and washed protected from light three times in warm 50% formamide, 2xSSC pH 7.2 for 5 min each at 42 degrees and three times in 2xSSC for 5 min. Cells were briefly washed in 2xSSC at room temperature and counterstain in 0.2mg/mL DAPI solution for 2 minutes at room temperature and washed twice in 2xSSC for 5 min each. Coverslips were dried and mounted in Vectashield and fixed with a minimal amount of nail varnish. Acquisitions were performed on LSM 880 and 780 confocal microscopes. After tile scan, hundreds of single nuclei positions were spotted by the use of DAPI channel in X, Y and Z. Three-dimensional acquisitions were then made for each of these positions for all channels and saved for processing.

Batch Alleles Investigation Tool “BAIT” was designed to run on Definiens software and measure FISH three-dimensional inter-probe distances automatically. Nucleus segmentation was performed in three-dimensions by use of the auto-threshold function of Definiens on DAPI staining. Nuclear boundaries were analysed to

detect and discard incomplete nuclei in X, Y and Z. Probes signals were determined as pixel intensity values within the nucleus. To comparison size and intensity of the signals, geometrical measurements were performed to set an object center “Seed” and clusters of spots “Allele” for and between each channel. Spot center distances were then calculated in three dimensios and exported as comma-separated values (.csv). Quality control was performed under Definiens to detect aberrations and artefacts. Statistics were performed using R. Violin distribution plots were made with Prism 8. Figures were made with Fiji standard deviation projection.

Comparison of FISH results with Hi-C maps

Numbers are given for the most densely populated square at highest resolution within the range of the loop coordinates targeted by our FISH probes. High-resolution Hi-C maps for control, STAG1-RNAi, and STAG2-RNAi at 5Kb nominal resolution were generated using “Juicertools”. Values indicate unique contact counts with KR (balanced) normalization, and no normalization applied.

Length distribution, CTCF occupancy, and CTCF orientation of loops

The lists of loops called by HiCCUPS on each of the three maps (STAG1 RNAi, STAG2 RNAi and control) were re-organized into four groups - loops that are specific to each of the three maps (STAG1 specific, STAG2 specific and control specific loops) and loops that appear in at least two maps (common loops).
The length of each loop was calculated as the distance between the midpoints of the loop anchors, and the length distributions of loops in each group were visualized with the empirical cumulative distribution function. To determine CTCF occupancy, we expanded loop anchors smaller than 15kb to 15kb and counted the number of anchors that overlap with at least one CTCF ChiP-seq peak using bedtools. We calculated fold enrichment of CTCF occupancy by comparing the counts with the average overlap of ten random translational controls with the same length distribution as the loop anchors. Orientation of CTCF motifs at loop anchors was identified using MotifFinder in Juicer, and the proportion of inward oriented (following the convergent rule) CTCF motifs was reported.

Aggregate peak analysis (APA) of loops

We performed aggregate peak analysis of hiccups-called loops using juicer_tools apa. The aggregate enrichment of our sets of looping peaks in contact matrices was visualized by plotting a cumulative stack of sub-matrices around detected loop coordinates. For a map with 10 Kb resolution we generated squares of 210 Kb x 210 Kb summing all putative loop peaks in a way that the resulting APA plot displays the total number of contacts which lie within the entire loop set at the center of the matrix within the aggregate pile-ups of their surroundings (Rao et al., 2014). All underlying matrices are KR-normalized.

Aggregate analysis of TADs

For the analysis and visualization of average TAD pileups we generated size-sorted and -classified lists of TADs and calculated histogram matrices around the centers of those TAD areas within narrow predefined size ranges. That way several sub-matrices of interaction around similarly size-classified TADs were added up in order to generate a global profile of one size-range for every Hi-C matrix (10 Kb resolution, coverage-normalization).

Hi-C insulation plots

Plots of the insulation scores were made based on insulation bedGraph files and TAD boundary coordinates generated by using the 'findTADsAndLoops.pl' script in the HOMER software package. This software scans relative contact matrices for locally dense regions of contacts or areas with an increased degree of intra-domain interactions relative to surrounding regions. Using a resolution of 3000, a window size of 15000, and the default maximum interaction distance (2MB), we generated a coordinate set of sites with maximal transition in contact orientation, i.e. sites with highest insulation. Average plots of insulation profiles in all samples were made from regions centered around the coordinates in the respective bedGraph files.

Simulations

Molecular dynamic simulations were performed using HOOMD-blue (Anderson et al., 2008; Glaser et al., 2015) using an approach similar to (Sanborn et al., 2015) with minor modifications. The region 17.25-19.25mb on chromosome 9 under wild type, STAG1-RNAi, and STAG2-RNAi conditions (*Figure 5B*) was simulated as a polymer of length 2000, each monomer representing 1kb of chromatin, for a total of 850000 times steps. A total of 480 replicated simulations were performed and aggregated into a single contact map. All simulations contained an average of three cohesin complexes actively extruding chromatin to form loops. The probability for one end of the extrusion complex to halt at a particular locus was derived from CTCF ChIP-seq data in HeLa cells (ENCODE phase 2, Broad Institute, file ENCFF000BAN), normalized to a probability between 0 and 1. Each halted end of an extrusion complex also had a 0.05% probability to continue sliding at each time step (halting lifetime of ~2000 time steps). In addition, extrusion complexes have had predefined average lifetime on chromatin, and changes in STAG1 and STAG2 levels were simulated by modulating the extrusion lifetime. Wild type extrusion complexes had an average lifetime of 5000 timesteps (0.02% chance of dissociating at each time step) while STAG1-RNAi had 5-fold higher lifetimes and STAG2-RNAi had 2-fold lower lifetimes.
Simplified simulations of short, medium, or long extrusion lifetimes (Figure 5A) were performed similarly. Chromatin was represented as polymers of length 1000 containing an average of three extrusion complexes. The two outermost CTCF binding sites had halting probabilities of 0.9 while the inner four binding sites had halting probabilities of 0.4. Halted extrusion ends had a 0.01% probability to continue sliding. Short, medium, and long extrusion lifetimes were modeled as complexes with lifetimes of 312.5, 625, and 10000 time steps respectively.

ChIP-seq peak calling and calculation of peak overlaps

Cohesin (SMC3) and CTCF ChIP-seq was performed as described in (Wendt et al., 2008). Peaks were called by the MACS algorithm version 1.4.2 (Zhang et al., 2008), using a P-value threshold of 1e-10 and by using sample and input read files. We identified sites of overlapping peaks between different conditions as well as between SMC3 and CTCF peaks using the MULTOVL software (Aszódi, 2012). We applied an inclusive type of overlap display (“union”), in which coordinates of overlapping peaks are merged into one common genomic site.

Hi-C library preparation

We generated a total of 19 *in situ* Hi-C libraries from our RNAi experiments. Libraries generated by using Mbo I enzyme were done as described in Rao et al., 2014 without modification. In brief, the *in situ* Hi-C protocol involves crosslinking cells with formaldehyde, permeabilizing nuclei with detergent, digesting DNA overnight using a 4-cutter restriction enzyme, filling in 5’-overhangs while incorporating a biotinylated nucleotide, ligating newly blunted ends together, shearing DNA, capturing biotinylated ligation junctions with

streptavidin beads, and analyzing the resulting fragments with paired-end sequencing. All the libraries generated with the 6 bp cutter Hind III were performed as in (Wutz et al., 2017).

Hi-C Data Processing

All Hi-C libraries were sequenced with 150 bp paired-end reads on an Illumina HiSeqX. All resulting data was processed using "Juicer tool" (Durand et al., 2016b; Rao et al., 2014) and aligned against the hg19 reference genome. All contact matrices used for analysis were Knight-Ruiz or Vanilla-Coverage normalized with Juicer.

Loops were annotated in our RNAi experiments using HiCCUPS (Durand et al., 2016a; Rao et al., 2014). Default parameters as described in (Durand et al., 2016a; Rao et al., 2014) were used to call loops at 5 kb and 10 kb resolutions and merged as described in Rao et al., 2014 with the following exceptions. All loop calls on chromosomes 4 across all maps was called using Vanilla-Coverage normalization while the remaining loops calls on all other chromosomes were done using Knight-Ruiz normalization. We observed that karyotype abnormalities in the Kyoto Hela cell line were annotated as loops across all experimental conditions so all peak calls that displayed an observed/expected bottom left enrichment of >4.5 were removed. Empirically, this threshold removed peak annotations that were the result of the karyotypic abnormalities in the Kyoto Hela.

Domains were annotated in our RNAi experiments using Arrowhead (Durand et al., 2016a; Rao et al., 2014). Domains were called at 5kb and 10kb resolutions using default parameters and merged.

Data Availability

The Hi-C and ChIP-seq sequencing data from this publication have been deposited in the Gene Expression Omnibus (GEO) under the accession number

Acknowledgements

Research in the laboratory of J-MP was supported by Boehringer Ingelheim, the Austrian Science Fund (FWF special research program SFB F34 "Chromosome Dynamics" and Wittgenstein award Z196-B20), the Austrian Research Promotion Agency (Headquarter grants FFG-834223 and FFG-852936, Laura Bassi Centre for Optimized Structural Studies grant FFG-840283), the Vienna Science and Technology Fund (WWTF LS09-13), the European Community's Seventh Frame- work Programme (FP7/2007-2013) under grant agreement 241548 (MitoSys) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme GA No 693949. The work in the lab of KM has been supported by EPIC-XS, project number 823839, funded by the Horizon 2020 program of the European Union, and the Austrian Science Fund by ERA-CAPS I 3686 International Project. Research in the laboratory of ELA was supported by NIH (5UM1HG009375-03 and 5U01HL130010-05) and NSF (PHY-1427654).

Author contributions

GW and JMP conceived the project. GW, RL, BP, MI and PVL designed experiments and interpreted data together with RS, CV and XH. Cell lines were generated by RL (EGFP-STAG1, STAG1-EGFP, STAG2-EGFP, EGFP-STAG1 FLAG-RFP-STAG2) and WT (STAG1-AID-GFP and STAG2-AID-GFP). RL, KN and GW performed RNAi and immunoblotting and microscopy experiments. WT performed ChIP-seq

experiments. GW and BSH generated the Hi-C libraries with help from SS. RS, CV and BSH performed bioinformatic analyses of Hi-C data with input from GW. AS performed modeling. RS analyzed ChIP-seq data. ELA, PF, KM and JMP supervised the project. IFD, GW, ELA and JMP wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Anderson JA, Lorenz CD, Travesset A. 2008. General purpose molecular dynamics simulations fully implemented on graphics processing units. *J Comput Phys*. doi:10.1016/j.jcp.2008.01.047

Aszódi A. 2012. MULTOVL: Fast multiple overlaps of genomic regions. *Bioinformatics*. doi:10.1093/bioinformatics/bts607

Ben-Shahar TR, Heeger S, Lehane C, East P, Flynn H, Skehel M, Uhlmann F. 2008. Eco1-dependent cohesin acetylation during establishment of sister chromatid cohesion. *Science (80-)*. doi:10.1126/science.1157774

Burkhardt S, Borsos M, Szydlowska A, Godwin J, Williams SA, Cohen PE, Hirota T, Saitou M, Tachibana-Konwalski K. 2016. Chromosome Cohesion Established by Rec8-Cohesin in Fetal Oocytes is Maintained without Detectable Turnover in Oocytes Arrested for Months in Mice. *Curr Biol*. doi:10.1016/j.cub.2015.12.073

Busslinger GA, Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters J-M. 2017a. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**:503–507. doi:10.1038/nature22063

Busslinger GA, Stocsits RR, Van Der Lelij P, Axelsson E, Tedeschi A, Galjart N, Peters JM. 2017b. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*. doi:10.1038/nature22063

Cai Y, Hossain MJ, Hériché JK, Politi AZ, Walther N, Koch B, Wachsmuth M, Nijmeijer B, Kueblbeck M, Martinic-Kavur M, Ladurner R, Alexander S, Peters JM, Ellenberg J. 2018. Experimental and computational framework for a dynamic protein atlas of human cell division. *Nature*. doi:10.1038/s41586-018-0518-z

Casa V, Moronta Gines MM, Gusmao EG, Slotman JA, Zirkel A, Josipovic N, Oole E, van IJcken WFJ, Houtsmuller AB, Papantonis A, Wendt KS. 2019. Redundant and specific roles of cohesin STAG subunits in chromatin looping and transcription control. *bioRxiv*. doi: https://doi.org/10.1101/642959

Chan KL, Gligoris T, Upcher W, Kato Y, Shirahige K, Nasmyth K, Beckouët F. 2013. Pds5 promotes and protects cohesin acetylation. *Proc Natl Acad Sci U S A*. doi:10.1073/pnas.1306900110

Chan KL, Roig MB, Hu B, Beckouët F, Metson J, Nasmyth K. 2012. Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell*. doi:10.1016/j.cell.2012.07.028

Cuadrado A, Giménez-Llorente D, Kojic A, Rodríguez-Corsino M, Cuartero Y, Martín-Serrano G, Gómez-López G, Marti-Renom MA, Losada A. 2019. Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and Polycomb Domains in Embryonic Stem Cells. *Cell Rep*.

doi:10.1016/j.celrep.2019.05.078

Dauban L, Montagne R, Thierry A, Lazar-Stefanita L, Gadal O, Cournac A, Koszul R, Beckouët F.2019.A major role for Eco1 in regulating cohesin-mediated mitotic chromosome folding. *bioRxiv*. doi: https://doi.org/10.1101/589101

de Wit E, Vos ESM, Holwerda SJB, Valdes-Quezada C, Verstegen MJAM, Teunissen H, Splinter E, Wijchers PJ, Krijger PHL, de Laat W. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell*. doi:10.1016/j.molcel.2015.09.023

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*. doi:10.1038/nature11082

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016a. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst*. doi:10.1016/j.cels.2015.07.012

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016b. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst*. doi:10.1016/j.cels.2016.07.002

Elbashir SM, Martinez J, Patkaniowska A, Lendeckel W, Tuschl T. 2001. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J*. doi:10.1093/emboj/20.23.6877

Flyamer IM, Gassler J, Imakaev M, Brandão HB, Ulianov S V., Abdennur N, Razin S V., Mirny LA, Tachibana-Konwalski K. 2017. Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature*. doi:10.1038/nature21711

Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*. doi:10.1016/j.celrep.2016.04.085

Ganji M, Shaltiel IA, Bisht S, Kim E, Kalichava A, Haering CH, Dekker C. 2018. Real-time imaging of DNA loop extrusion by condensin. *Science (80-)*. doi:10.1126/science.aar7831

Gassler J, Brandão HB, Imakaev M, Flyamer IM, Ladstätter S, Bickmore WA, Peters J, Mirny LA, Tachibana K. 2017. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J*. doi:10.15252/embj.201798083

Gerlich D, Koch B, Dupeux F, Peters JM, Ellenberg J. 2006. Live-Cell Imaging Reveals a Stable Cohesin-Chromatin Interaction after but Not before DNA Replication. *Curr Biol*. doi:10.1016/j.cub.2006.06.068

Gibcus JH, Samejima K, Goloborodko A, Samejima I, Naumova N, Nuebler J, Kanemaki MT, Xie L, Paulson JR, Earnshaw WC, Mirny LA, Dekker J. 2018. A pathway for mitotic chromosome formation. *Science (80-)*. doi:10.1126/science.aao6135

Glaser J, Nguyen TD, Anderson JA, Lui P, Spiga F, Millan JA, Morse DC, Glotzer SC. 2015. Strong scaling of general-purpose molecular dynamics simulations on GPUs. *Comput Phys Commun*. doi:10.1016/j.cpc.2015.02.028

Haarhuis JHI, van der Weide RH, Blomen VA, Yáñez-Cuna JO, Amendola M, van Ruiten MS, Krijger PHL, Teunissen H, Medema RH, van Steensel B, Brummelkamp TR, de Wit E, Rowland BD. 2017. The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell*. doi:10.1016/j.cell.2017.04.013

Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M. 2009. Cohesins form chromosomal cis-interactions at the developmentally regulated IFNG locus. *Nature*. doi:10.1038/nature08079

Haering CH, Farcas AM, Arumugam P, Metson J, Nasmyth K. 2008. The cohesin ring concatenates sister DNA molecules. *Nature*. doi:10.1038/nature07098

Hara K, Zheng G, Qu Q, Liu H, Ouyang Z, Chen Z, Tomchick DR, Yu H. 2014. Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat Struct Mol Biol*. doi:10.1038/nsmb.2880

Holzmann J, Politi AZ, Nagasaka K, Hantsche-Grininger M, Walther N, Koch B, Fuchs J, Dürnberger G, Tang W, Ladurner R, Stocsits RR, Busslinger GA, Novák B, Mechtler K, Davidson IF, Ellenberg J, Peters J-M. 2019. Absolute quantification of cohesin, CTCF and their regulators in human cells. *Elife*. doi:10.7554/elife.46269

Kawasumi R, Abe T, Arakawa H, Garre M, Hirota K, Branzei D. 2017. ESCO1/2's roles in chromosome structure and interphase chromatin organization. *Genes Dev*. doi:10.1101/gad.306084.117

Kojic A, Cuadrado A, De Koninck M, Giménez-Llorente D, Rodríguez-Corsino M, Gómez-López G, Le Dily F, Marti-Renom MA, Losada A. 2018. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome organization. *Nat Struct Mol Biol*. doi:10.1038/s41594-018-0070-4

Kueng S, Hegemann B, Peters BH, Lipp JJ, Schleiffer A, Mechtler K, Peters JM. 2006. Wapl Controls the Dynamic Association of Cohesin with Chromatin. *Cell*. doi:10.1016/j.cell.2006.09.040

Ladurner R, Bhaskara V, Huis In 'T Veld PJ, Davidson IF, Kreidl E, Petzold G, Peters JM. 2014. Cohesin's ATPase activity couples cohesin loading onto DNA with Smc3 acetylation. *Curr Biol*. doi:10.1016/j.cub.2014.08.011

Ladurner R, Kreidl E, Ivanov MP, Ekker H, Idarraga-Amado MH, Busslinger GA, Wutz G, Cisneros DA, Peters J. 2016. Sororin actively maintains sister chromatid cohesion. *EMBO J*. doi:10.15252/embj.201592532

Lafont AL, Song J, Rankin S. 2010. Sororin cooperates with the acetyltransferase Eco2 to ensure DNA replication-dependent sister chromatid cohesion. *Proc Natl Acad Sci*. doi:10.1073/pnas.1011069107

Landry JJM, Pyl PT, Rausch T, Zichner T, Tekkedil MM, Stütz AM, Jauch A, Aiyar RS, Pau G, Delhomme N, Gagneur J, Korbel JO, Huber W, Steinmetz LM. 2013. The genomic and transcriptomic landscape of a hela cell line. *G3 Genes, Genomes, Genet*. doi:10.1534/g3.113.005777

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*. doi:10.1038/nature12912

Leiserson MDM, Vandin F, Wu HT, Dobson JR, Eldridge J V., Thomas JL, Papoutsaki A, Kim Y, Niu B, McLellan M, Lawrence MS, Gonzalez-Perez A, Tamborero D, Cheng Y, Ryslik GA, Lopez-Bigas N, Getz G, Ding L, Raphael BJ. 2015. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat Genet*. doi:10.1038/ng.3168

Lin SG, Ba Z, Alt FW, Zhang Y. 2018. RAG Chromatin Scanning During V(D)J Recombination and Chromatin Loop Extrusion are Related ProcessesAdvances in Immunology. doi:10.1016/bs.ai.2018.07.001

Losada A, Yokochi T, Kobayashi R, Hirano T. 2000. Identification and characterization of SA/Scc3p

subunits in the Xenopus and human cohesin complexes. *J Cell Biol*. doi:10.1083/jcb.150.3.405

Merkenschlager M, Nora EP. 2016. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annu Rev Genomics Hum Genet*. doi:10.1146/annurev-genom-083115-022339

Michalska AE. 2007. Isolation and Propagation of Mouse Embryonic Fibroblasts and Preparation of Mouse Embryonic Feeder Layer CellsCurrent Protocols in Stem Cell Biology. doi:10.1002/9780470151808.sc01c03s3

Minamino M, Ishibashi M, Nakato R, Akiyama K, Tanaka H, Kato Y, Negishi L, Hirota T, Sutani T, Bando M, Shirahige K. 2015. Esco1 Acetylates Cohesin via a Mechanism Different from That of Esco2. *Curr Biol*. doi:10.1016/j.cub.2015.05.017

Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. 2013. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*. doi:10.1038/nature12593

Nagano T, Lubling Y, Várnai C, Dudley C, Leung W, Baran Y, Mendelson Cohen N, Wingett S, Fraser P, Tanay A. 2017. Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature*. doi:10.1038/nature23001

Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, Murrell A. 2009. Cohesin is required for higher-order chromatin conformation at the imprinted IGF2-H19 locus. *PLoS Genet*. doi:10.1371/journal.pgen.1000739

Nishiyama T, Ladurner R, Schmitz J, Kreidl E, Schleiffer A, Bhaskara V, Bando M, Shirahige K, Hyman AA, Mechtler K, Peters JM. 2010. Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell*. doi:10.1016/j.cell.2010.10.031

Nora EP, Goloborodko A, Valton AL, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell*. doi:10.1016/j.cell.2017.05.004

Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, Van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. doi:10.1038/nature11049

Ouyang Z, Zheng G, Tomchick DR, Luo X, Yu H. 2016. Structural Basis and IP6 Requirement for Pds5-Dependent Cohesin Dynamics. *Mol Cell*. doi:10.1016/j.molcel.2016.02.033

Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, Cobb BS, Yokomori K, Dillon N, Aragon L, Fisher AG, Merkenschlager M. 2008. Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. *Cell*. doi:10.1016/j.cell.2008.01.011

Poser I, Sarov M, Hutchins JRA, Hériché JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW, Pelletier L, Kittler R, Hua S, Naumann R, Augsburg M, Sykora MM, Hofemeister H, Zhang Y, Nasmyth K, White KP, Dietzel S, Mechtler K, Durbin R, Stewart AF, Peters JM, Buchholz F, Hyman AA. 2008. BAC TransgeneOmics: A high-throughput method for exploration of protein function in mammals. *Nat Methods*. doi:10.1038/nmeth.1199

Rankin S, Ayad NG, Kirschner MW. 2005. Sororin, a substrate of the anaphase- promoting complex, is required for sister chromatid cohesion in vertebrates. *Mol Cell*. doi:10.1016/j.molcel.2005.03.017

Rao SSP, Huang SC, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon KR, Sanborn AL,

Johnstone SE, Bascom GD, Bochkov ID, Huang X, Shamim MS, Shin J, Turner D, Ye Z, Omer AD, Robinson JT, Schlick T, Bernstein BE, Casellas R, Lander ES, Aiden EL. 2017. Cohesin Loss Eliminates All Loop Domains. *Cell*. doi:10.1016/j.cell.2017.09.026

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. doi:10.1016/j.cell.2014.11.021

Remeseiro S, Cuadrado A, Gómez-Lãpez G, Pisano DG, Losada A. 2012. A unique role of cohesin-SA1 in gene regulation and development. *EMBO J*. doi:10.1038/emboj.2012.60

Roig MB, Löwe J, Chan KL, Beckouët F, Metson J, Nasmyth K. 2014. Structure and function of cohesin's Scc3/SA regulatory subunit. *FEBS Lett*. doi:10.1016/j.febslet.2014.08.015

Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL. 2015. SI: Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci*. doi:10.1073/pnas.1518552112

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: An open-source platform for biological-image analysis. *Nat Methods*. doi:10.1038/nmeth.2019

Schmitz J, Watrin E, Lénárt P, Mechtler K, Peters JM. 2007. Sororin Is Required for Stable Binding of Cohesin to Chromatin and for Sister Chromatid Cohesion in Interphase. *Curr Biol*. doi:10.1016/j.cub.2007.02.029

Schwarzer W, Abdennur N, Goloborodko A, Pekowska A, Fudenberg G, Loe-Mie Y, Fonseca NA, Huber W, H Haering C, Mirny L, Spitz F. 2017. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. doi:10.1038/nature24281

Shintomi K, Hirano T. 2009. Releasing cohesin from chromosome arms in early mitosis: Opposing actions of Wapl-Pds5 and Sgo1. *Genes Dev*. doi:10.1101/gad.1844309

Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, Toretsky JA, Rosenberg SA, Shukla N, Ladanyi M, Samuels Y, James CD, Yu H, Kim JS, Waldman T. 2011. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science (80-)*. doi:10.1126/science.1203619

Strunnikov A V., Larionov VL, Koshland D. 1993. SMC1: An essential yeast gene encoding a putative head-rod-tail protein is required for nuclear division and defines a new ubiquitous protein family. *J Cell Biol*. doi:10.1083/jcb.123.6.1635

Sumara I, Vorlaufer E, Gieffers C, Peters BH, Peters JM. 2000. Characterization of vertebrate cohesin complexes and their regulation in prophase. *J Cell Biol*. doi:10.1083/jcb.151.4.749

Tachibana-Konwalski K, Godwin J, Van Der Weyden L, Champion L, Kudo NR, Adams DJ, Nasmyth K. 2010. Rec8-containing cohesin maintains bivalents without turnover during the growing phase of mouse oocytes. *Genes Dev*. doi:10.1101/gad.605910

Tedeschi A, Wutz G, Huet S, Jaritz M, Wuensche A, Schirghuber E, Davidson IF, Tang W, Cisneros DA, Bhaskara V, Nishiyama T, Vaziri A, Wutz A, Ellenberg J, Peters JM. 2013. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature*. doi:10.1038/nature12471

Ünal E, Heidinger-Pauli JM, Kim W, Guacci V, Onn I, Gygi SP, Koshland DE. 2008. A molecular determinant for the establishment of sister chromatid cohesion. *Science (80-)*. doi:10.1126/science.1157880

Vaur S, Feytout A, Vazquez S, Javerzat JP. 2012. Pds5 promotes cohesin acetylation and stable cohesin-chromosome interaction. *EMBO Rep*. doi:10.1038/embor.2012.72

Vian L, Pękowska A, Rao SSP, Kieffer-Kwon KR, Jung S, Baranello L, Huang SC, El Khattabi L, Dose M, Pruett N, Sanborn AL, Canela A, Maman Y, Oksanen A, Resch W, Li X, Lee B, Kovalchuk AL, Tang Z, Nelson S, Di Pierro M, Cheng RR, Machol I, St Hilaire BG, Durand NC, Shamim MS, Stamenova EK, Onuchic JN, Ruan Y, Nussenzweig A, Levens D, Aiden EL, Casellas R. 2018. The Energetics and Physiological Impact of Cohesin Extrusion. *Cell*. doi:10.1016/j.cell.2018.03.072

Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, Hadjur S. 2015. Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. *Cell Rep*. doi:10.1016/j.celrep.2015.02.004

Viny AD, Bowman RL, Liu Y, Lavallée VP, Eisman SE, Xiao W, Durham BH, Navitski A, Park J, Braunstein S, Alija B, Karzai A, Csete IS, Witkin M, Azizi E, Baslan T, Ott CJ, Pe'er D, Dekker J, Koche R, Levine RL.2019. Cohesin Members Stag1 and Stag2 Display Distinct Roles in Chromatin Accessibility and Topological Control of HSC Self-Renewal and Differentiation. *Cell Stem Cell.* doi: 10.1016/j.stem.2019.08.003

Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters J-M. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**:796–801. doi:10.1038/nature06634

Whelan G, Kreidl E, Wutz G, Egner A, Peters JM, Eichele G. 2012. Cohesin acetyltransferase Esco2 is a cell viability factor and is required for cohesion in pericentric heterochromatin. *EMBO J*. doi:10.1038/emboj.2011.381

Wutz G, Várnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ, Walther N, Koch B, Kueblbeck M, Ellenberg J, Zuber J, Fraser P, Peters J-M. 2017. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J*. doi:10.15252/embj

Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. doi:10.1186/1471-2105-13-134

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**:R137. doi:10.1186/gb-2008-9-9-r137

Supplemental figures

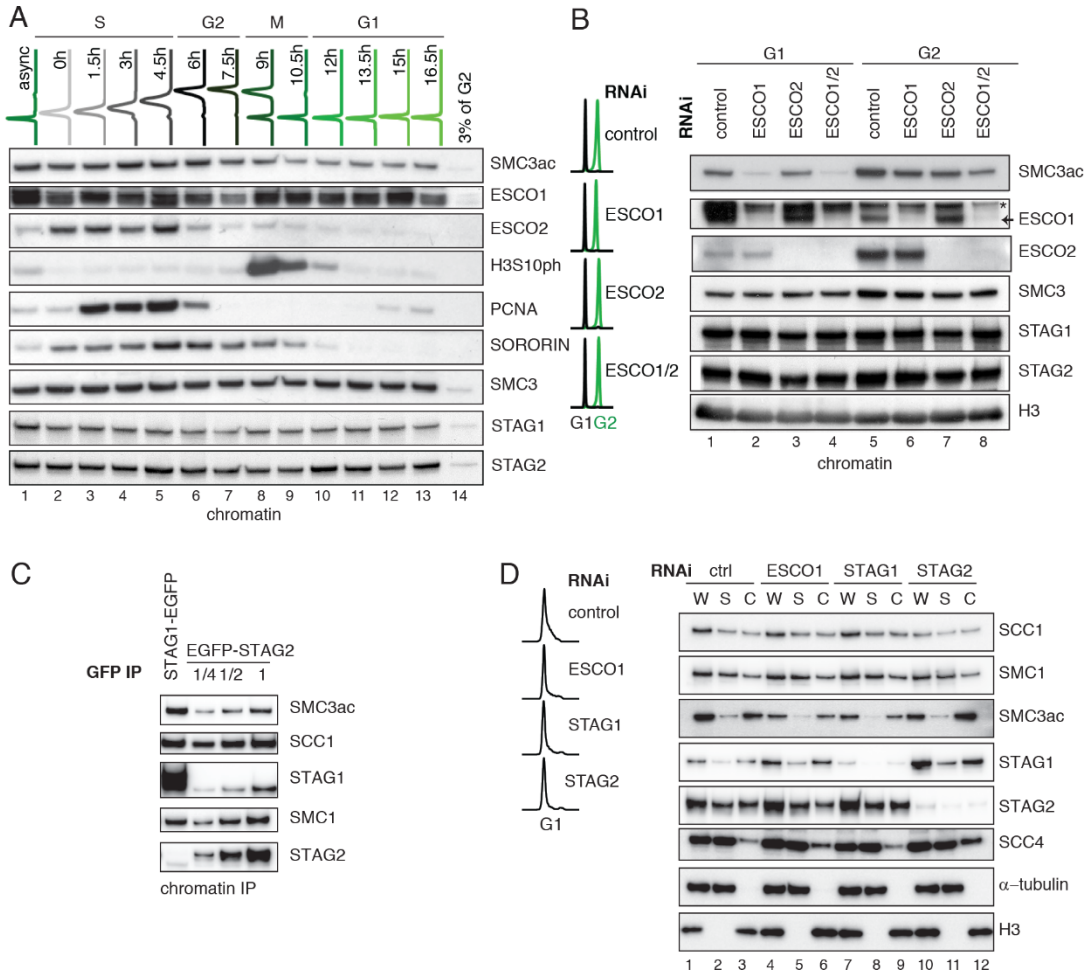
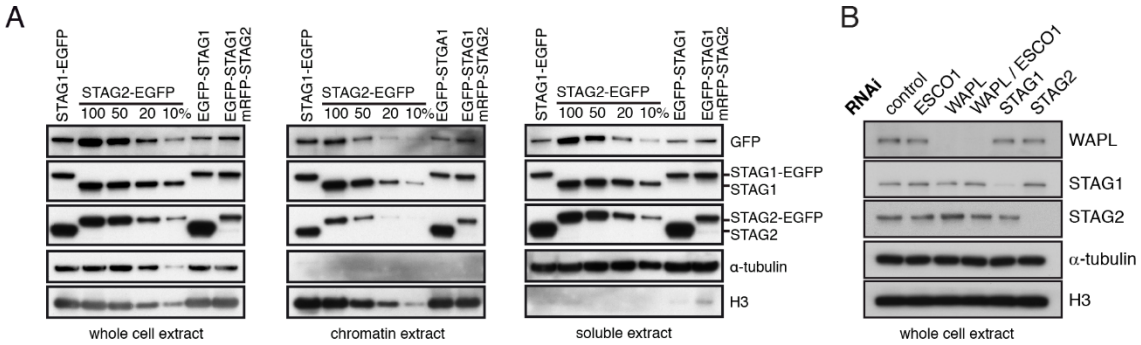
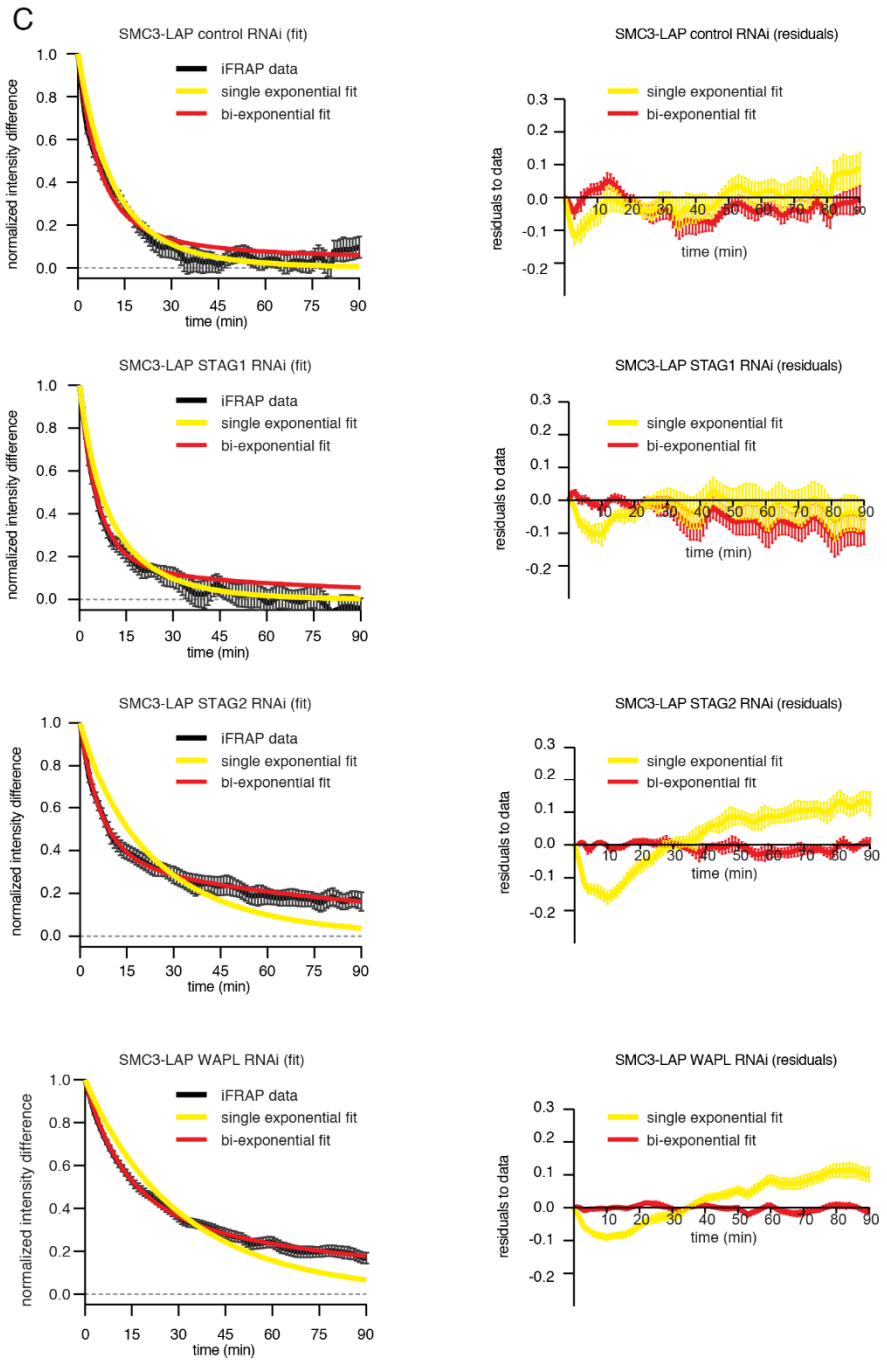


Figure S1. ESCO1 preferentially acetylates cohesinSTAG1 during G1 phase.

A. Flow cytometry and chromatin extract immunoblot analysis of cells at different times after double thymidine arrest/release. At the indicated time points, the chromatin pellet fraction was analyzed by immunoblotting using the indicated antibodies. Flow-cytometry profiles of different time points are shown on top.
B. Flow cytometry and chromatin extract immunoblot analysis of control, ESCO1, ESCO2 and ESCO1/2-depleted cells synchronized in G1 and G2. Immunoblotting was performed using the indicated antibodies. H3 was used as a loading control.
C. Immunoblot analysis of STAG1 and STAG2 immunoprecipitates obtained from chromatin extracts from equal numbers of EGFP-STAG1 and STAG2-EGFP HeLa cells. Immunoprecipitations were performed using anti-GFP antibodies.
D. Flow cytometry and immunoblot analysis of control, ESCO1, STAG1 and STAG2-depleted cells synchronized in G1. Immunoblotting of whole cell (W), supernatant (S) and chromatin(C) extracts was performed using the indicated antibodies. α-tubulin and H3 were used as loading controls.





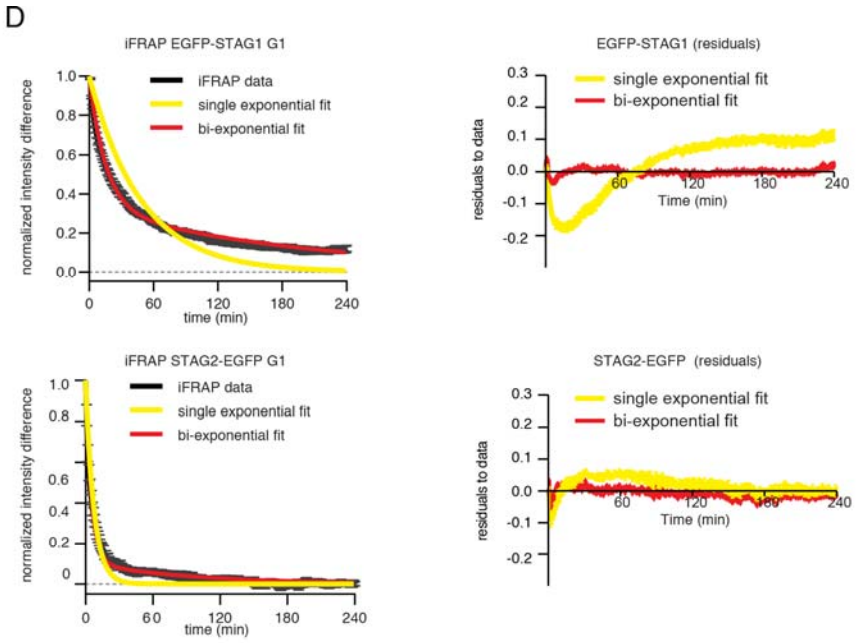


Figure S2. Characterisation of cell lines expressing tagged versions of STAG1 and STAG2 and iFRAP experiments

A. Immunoblot analysis of HeLa cell lines. Immunoblotting of whole cell, chromatin and soluble extracts was performed using the indicated antibodies. α-tubulin and H3 were used as loading controls.
B. Immunoblot analysis of whole cell extract from cells depleted of the proteins listed. Immunoblotting was performed using the indicated antibodies. α-tubulin and H3 were used as loading controls.
C. Curve fitting to experimental iFRAP curves obtained from SMC3-LAP in G1 phase by a single exponential and bi-exponential functions is shown on the left. Residuals representing the difference between the experimental- and simulated-value, of the fits from the left panel are plotted on the right.
D. Curve fitting to experimental iFRAP curves obtained from EGFP-STAG1 and STAG2-EGFP in G1 phase by a single exponential and bi-exponential functions is shown on the left. Residuals representing the difference between the experimental- and simulated-value, of the fits from the left panel are plotted on the right. and bi-exponential functions is shown on the left. Residuals representing the difference between the experimental- and simulated-value, of the fits from the left panel are plotted on the right.

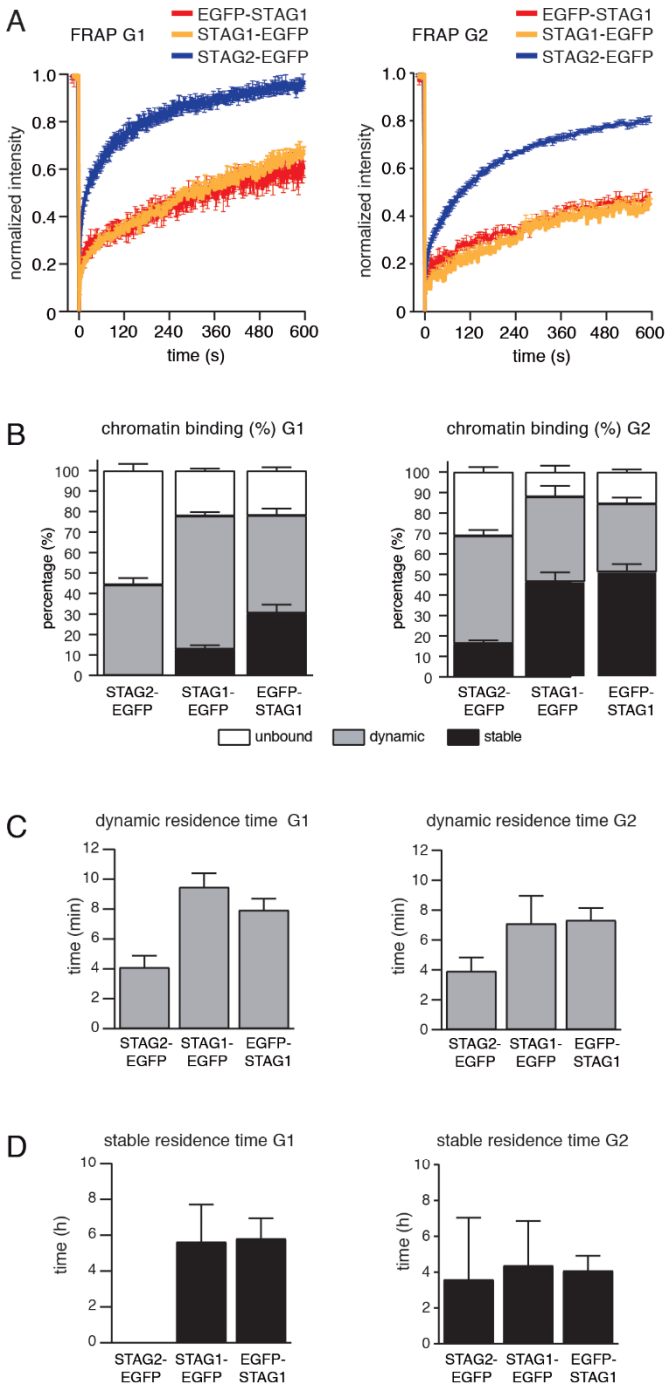


Figure S3. FRAP experiment confirms that subpopulation of cohesin^{STAG1} associates stably with chromatin during G1 phase

A. Graph depicting the mean normalized fluorescence intensity after photobleaching in EGFP-STAG1, STAG1-EGFP and STAG2-EGFP cells synchronized in G1 (left) and G2 (right). Error bars denote standard error of the mean (s.e.m.), n >= 6 cells per condition.
B. EGFP-STAG1, STAG1-EGFP and STAG2-EGFP distribution within the nucleus as measured by FRAP.
C. Quantification of the residence time of dynamically chromatin-bound STAG2-EGFP, STAG1-EGFP and EGFP-STAG1 in cells synchronized in G1 (left) and G2 (right).
D. Quantification of the residence time of stably chromatin-bound STAG2-EGFP, STAG1-EGFP and EGFP-STAG1 in cells synchronized in G1 (left) and G2 (right).

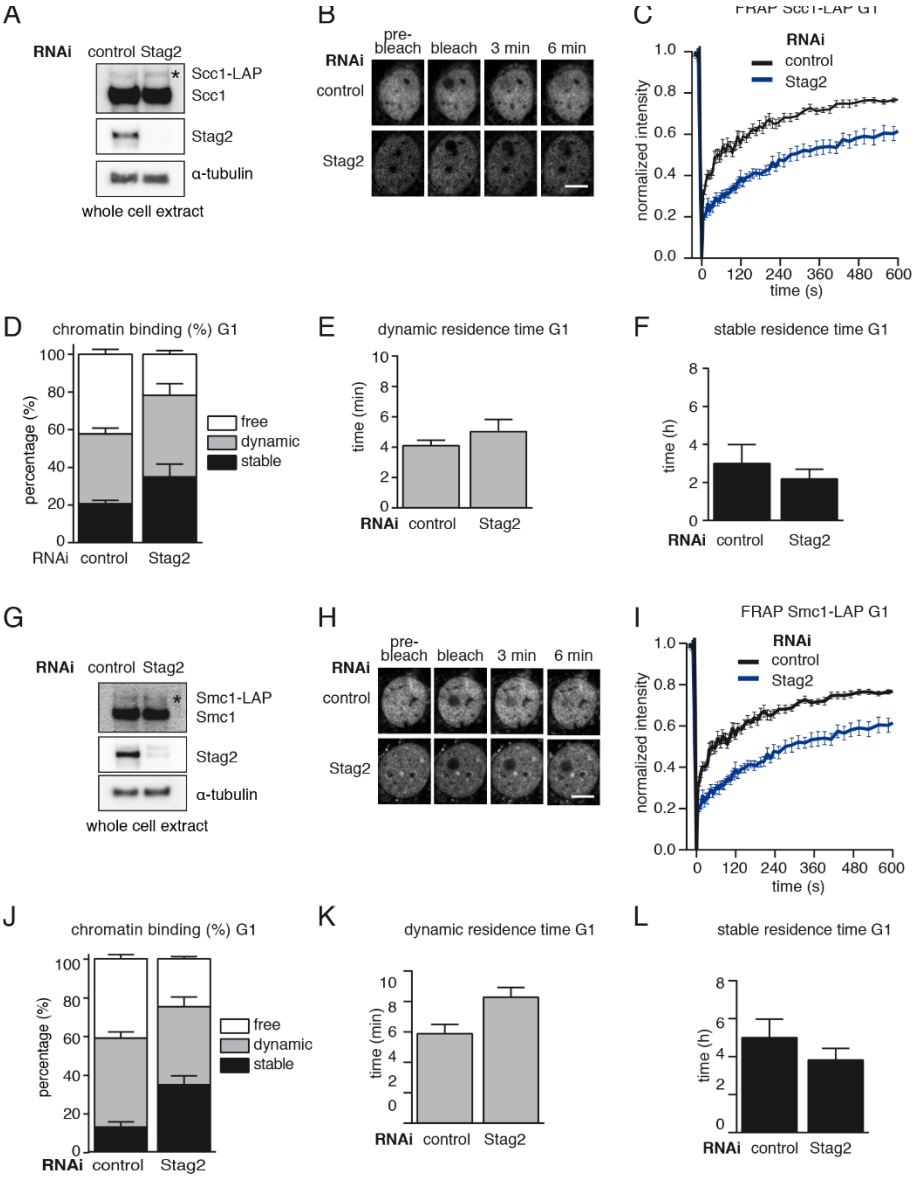


Figure S4. A subpopulation of cohesin^{STAG1} associates stably with chromatin in quiescent MEFs

A. Immunoblot analysis of whole cell extracts from control- and Stag2-depleted Scc1-LAP iMEFs cells synchronized in G0/1. Immunoblotting was performed using the indicated antibodies. The asterisk denotes the Scc1-LAP band. α-tubulin was used as a loading control.
B. Images of fluorescence recovery after photobleaching (FRAP) in Scc1-LAP cells synchronized in G0/G1 and depleted of Stag2 by RNAi. Scale bar, 10 μm.
C. Graph depicting the mean normalized fluorescence intensity after photobleaching in control- and Stag2-depleted Scc1-LAP cells synchronized in G1. Error bars denote standard error of the mean (s.e.m.), n = 10 cells per condition.
D. Scc1-LAP distribution within the nucleus of control- and Stag2-depleted G0/G1 cells, as measured by FRAP.
E. Quantification of the dynamically chromatin-bound residence time of Scc1-LAP in control- and Stag2-depleted cells synchronized in G1.
F. Quantification of the stably chromatin-bound residence time of Scc1-LAP in control- and Stag2-depleted cells synchronized in G1.
G. Immunoblot analysis of whole cell extracts from control- and Stag2-depleted Smc1-LAP iMEFs cells synchronized in G0/1. Immunoblotting was performed using the indicated antibodies. The asterisk denotes the Scc1-LAP band. α-tubulin was used as a loading control.
H. Images of fluorescence recovery after photobleaching (FRAP) in Smc1-LAP cells synchronized in G0/G1 and depleted of Stag2 by RNAi. Scale bar, 10 μm.
I. Graph depicting the mean normalized fluorescence intensity after photobleaching in control- and Stag2-depleted Smc1-LAP cells synchronized in G1. Error bars denote standard error of the mean (s.e.m.), n = 10 cells per condition.
J. Smc1-LAP distribution within the nucleus of control- and Stag2-depleted G0/G1 cells, as measured by FRAP.
K. Quantification of the dynamically chromatin-bound residence time of Smc1-LAP in control- and Stag2-depleted cells synchronized in G1.
L. Quantification of the stably chromatin-bound residence time of Smc1-LAP in control- and Stag2-depleted cells synchronized in G1.

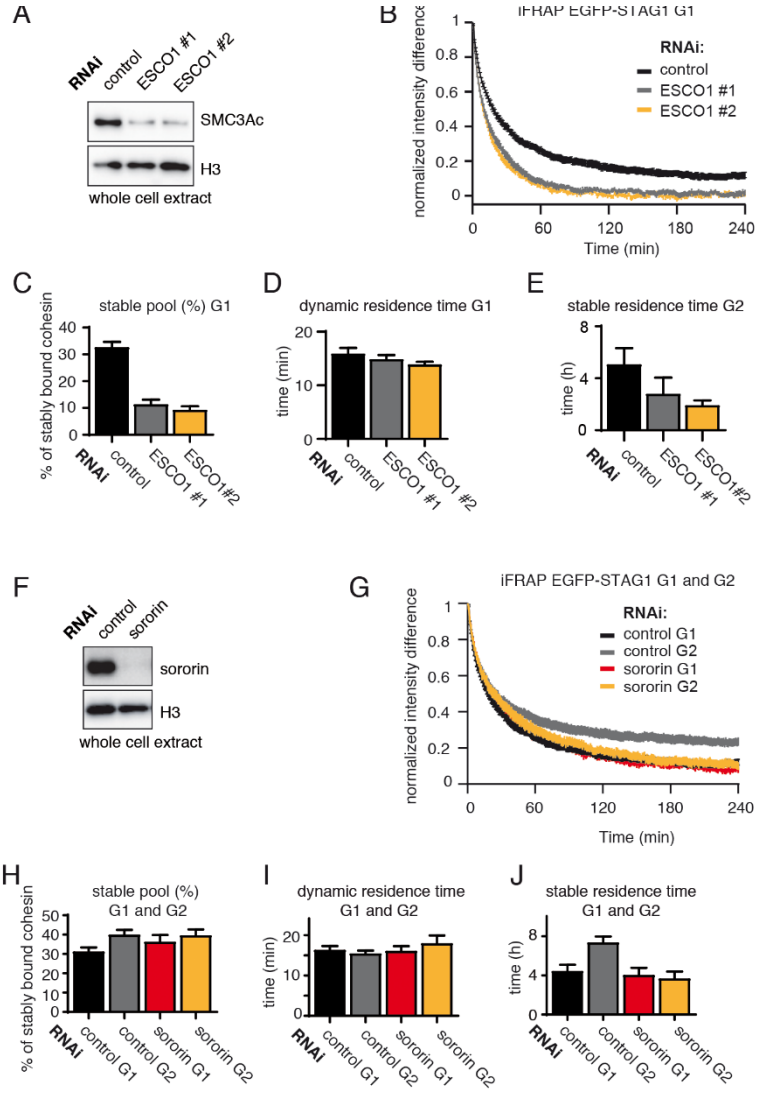


Figure S5. The long chromatin residence time of cohesin^{STAG1} in G1 depends on ESCO1

A. Immunoblot analysis of whole cell extracts from control- and ESCO1-depleted HeLa cells synchronized in G1. Two different ESCO1-specific siRNA oligos were used. Immunoblotting was performed using the indicated antibodies. H3 was used as a loading control.
B. Graph depicting the mean normalized difference in EGFP-STAG1 fluorescence intensity between the bleached and unbleached regions following iFRAP in cells depleted of ESCO1 protein by two different siRNAs. Error bars denote standard error of the mean (s.e.m.), n = 10 cells per condition.
C. Quantification of the fraction of nuclear STAG1-EGFP that was stably chromatin-bound in cells depleted for ESCO1 by RNAi.
D. Quantification of the dynamically chromatin-bound residence time of EGFP-STAG1 in cells depleted for ESCO1 protein by RNAi.
E. Quantification of the stably chromatin-bound residence time of EGFP-STAG1 in cells depleted for ESCO1 protein by RNAi.
F. Immunoblot analysis of whole cell extracts from control- and sororin-depleted HeLa cells. Immunoblotting was performed using the indicated antibodies. H3 was used as a loading control.
G. Graph depicting the mean normalized difference in EGFP-STAG1 fluorescence intensity between the bleached and unbleached regions following iFRAP in cells synchronized in G1 or G2 and depleted for sororin by RNAi. Error bars denote standard error of the mean (s.e.m.), n = 10 cells per condition.
H. Quantification of the fraction of nuclear STAG1-EGFP that was stably chromatin-bound in G1 or G2 cells depleted for sororin by RNAi.
I. Quantification of the dynamically chromatin-bound residence time of STAG1-EGFP in G1 or G2 cells depleted for sororin by RNAi.
J. Quantification of the stably chromatin-bound residence time of STAG1-EGFP in G1 or G2 cells depleted for sororin by RNAi.

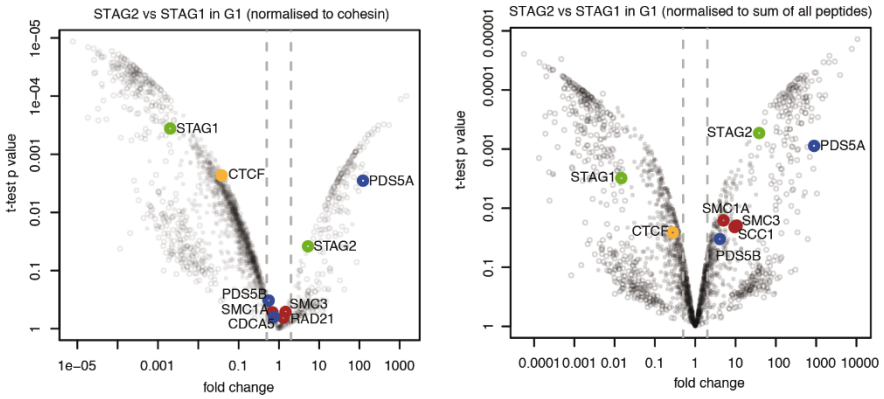


Figure S6. CTCF is enriched in cohesin^{STAG1} immunoprecipitates

Volcano plots of label-free qMS data, representing protein abundance in STAG2 immunoprecipitates relative to protein abundance in STAG1 immunoprecipitates. Area to the left of the grey dashed line denotes proteins significantly over-represented in STAG1 immunoprecipitates relative to STAG2 immunoprecipitates; vice versa is shown on the right. Left panel shows normalization relative to cohesin peptides and right panel shows normalization relative to the sum of all peptides.

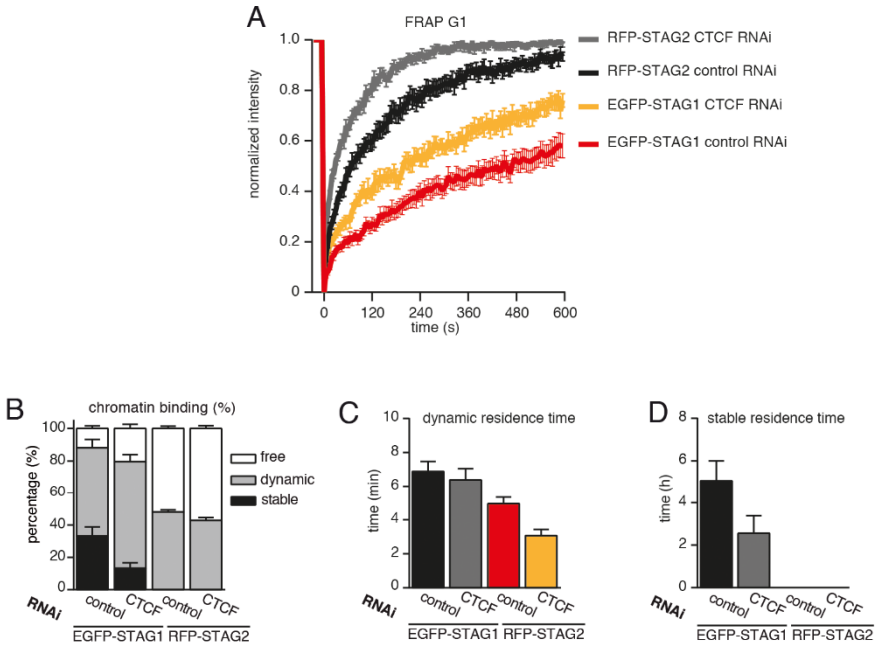


Figure S7. CTCF prolongs the residence time of cohesin on chromatin

A. Graph depicting the mean normalized RFP-STAG2 and EGFP-STAG1 fluorescence intensity after photobleaching in a clonal cell line that expressed both proteins tagged homozygously at their endogenous loci. Control- and CTCF-depleted cells were synchronized in G1. Error bars denote standard error of the mean (s.e.m.), n = 17 cells per condition.
B. Quantification of the fraction of nuclear RFP-STAG2 and EGFP-STAG1 that was stably chromatin-bound in cells synchronized in G1 and depleted of CTCF by RNAi.
C. Quantification of the residence time of dynamically chromatin-bound RFP-STAG2 and EGFP-STAG1 in cells synchronized in G1 and depleted of CTCF by RNAi.
D. Quantification of the residence time of stably chromatin-bound RFP-STAG2 and EGFP-STAG1 in cells synchronized in G1 and depleted of CTCF by RNAi.

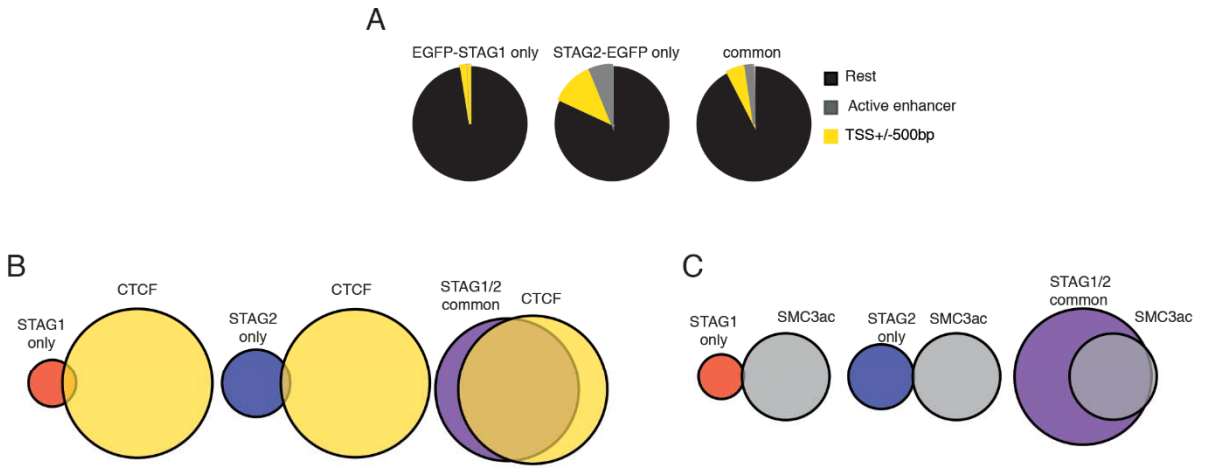


Figure S8. ChIP-seq analysis of the genomic distribution of STAG1, STAG2 and SMC3ac

A. Pie charts showing the distribution of STAG1-only (left), STAG2-only (middle) and STAG1/2 common ChIP-seq sites (right) in G1 relative to TSS and active enhancers.
B Pie charts showing the distribution of STAG1-only (left), STAG2-only (middle) and STAG1/2 common ChIP-seq sites (right) relative to CTCF ChIP-seq sites.
C. Pie charts showing the distribution of STAG1-only (left), STAG2-only (middle) and STAG1/2 common ChIP-seq sites (right) relative to SMC3ac ChIP-seq sites.

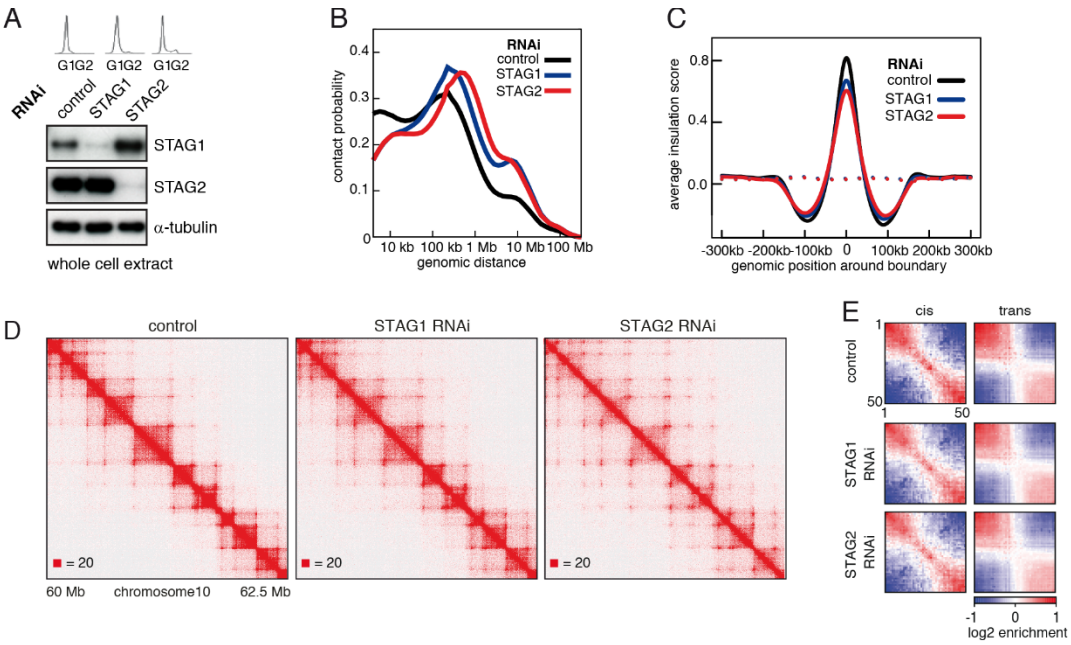
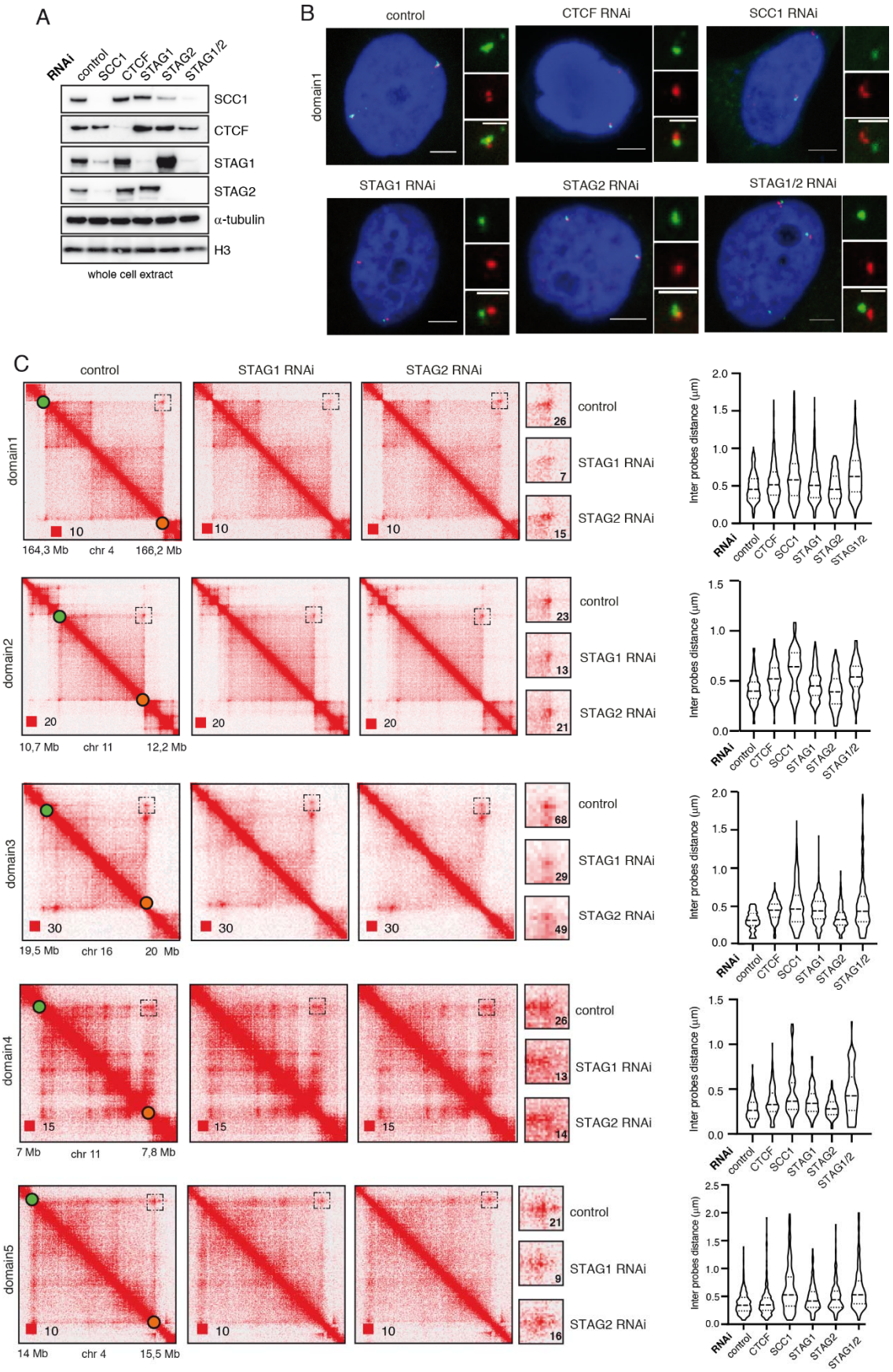


Figure S9. Cohesin^{STAG1} generates long chromatin interactions

A. Flow cytometry and whole cell extract immunoblot analysis of cells synchronized in G1 and depleted of the proteins indicated. α-tubulin was used as a loading control.
B. Intra-chromosomal contact frequency distribution as a function of genomic distance using logarithmically increasing genomic distance bins, for control-, STAG1- and STAG2-depleted HeLa cells.
C. Average insulation score around TAD boundaries in control-, STAG1- and STAG2-depleted HeLa cells synchronised in G1. Dashed lines indicate the average insulation score between random positions.
D. Coverage-corrected Hi-C matrices (chr 10: 60 – 62.5 Mb) from control, STAG1 and STAG2-depleted cells.
E. Long range (>2 Mb) intra-chromosomal (cis) and inter-chromosomal (trans) contact enrichment between bins with varying compartment signal strength from most B-like (1) to most A-like (50).



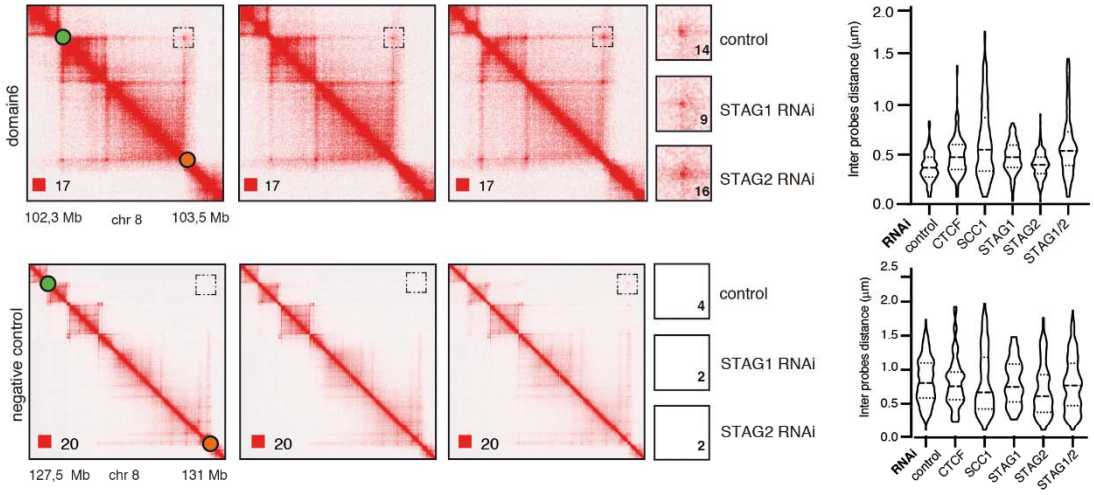


Figure S10. STAG1 depletion reduces loop signal more than STAG2 depletion

A. Immunoblot analysis of whole cell extracts from HeLa cells depleted for proteins as indicated using RNAi. Immunoblotting was performed using the indicated antibodies. α-tubulin and H3 were used as loading controls.
B. Representative FISH images from HeLa cells depleted of the indicated proteins. Scale bar, 10 µm.
C. Coverage-corrected Hi-C contact matrices of loci used in FISH analysis from control-, STAG1- and STAG2-depleted HeLa cells are shown on the left. The matrices were plotted using Juicebox. Boxed regions (also shown at higher magnification in the middle panels) represent the loop loci used in the FISH analysis. The numbers denote the most densely populated square at highest resolution within the range of the loop coordinates targeted by FISH probes. The right panel shows violin plots of the FISH 3D distance measurements between the loci boxed on the Hi-C matrices. The dashed lines show the median and interquartile range of between 100 and 500 cells for each condition.

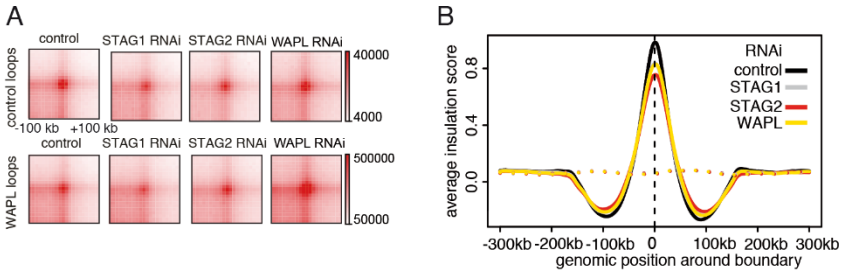


Figure S11. Cohesin^{STAG1} forms longer loops than cohesin^{STAG2}

A. Total contact counts around loops longer than 150 kb in control-, STAG1-, STAG2- and WAPL-depleted HeLa cells. Loops identified by HiCCUPS in control and WAPL-depleted cells.
B. Average insulation score around TAD boundaries in control-, STAG1-, STAG2- and WAPL-depleted HeLa cells synchronised in G1. Dashed lines indicate the average insulation score between random positions.

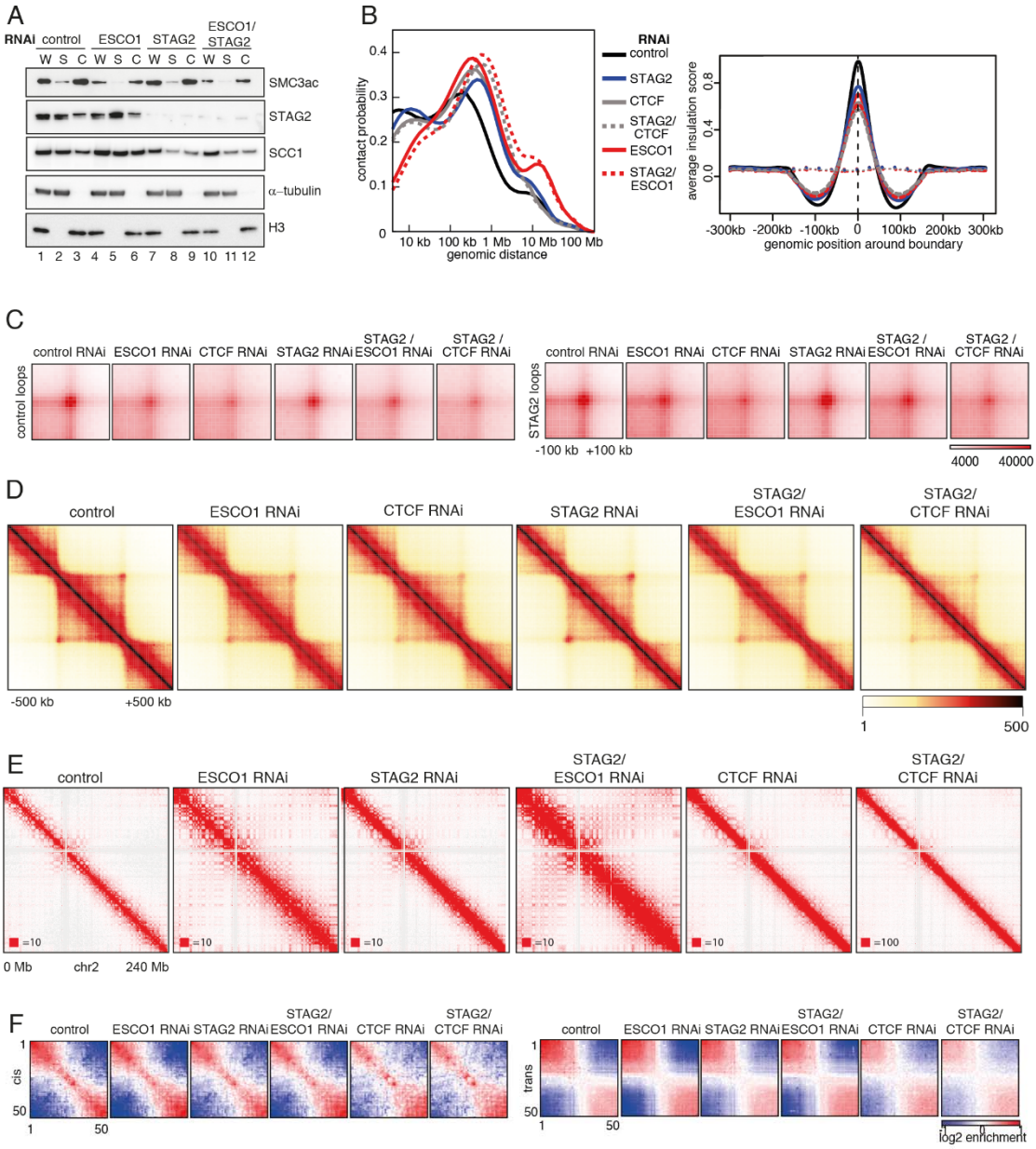


Figure S12. The Eefect of ESCO1 and CTCF depletion on chromatin organization

A. Immunoblot analysis of whole cell (W), cytoplasmic (S) and chromatin (C) extracts from control-, ESCO1-, STAG2- and ESCO1/STAG2-depleted HeLa cells synchronized in G1. Immunoblotting was performed using the indicated antibodies. ⍰-tubulin and H3 were used as loading control. α-tubulin and H3 were used as loading controls.
B. Left panel: intra-chromosomal contact frequency distribution as a function of genomic distance using logarithmically increasing genomic distance bins for control-, STAG2-, STAG2/CTCF-, ESCO1- and STAG2/ESCO1-depleted HeLa cells. Right panel: average insulation score around TAD boundaries. Dashed lines indicate the average insulation score between random positions.
C. Total contact counts around control-specific loops (> 150 kb) and STAG2-specific loops (> 150 kb) in control-, ESCO1-, CTCF-, STAG2-, STAG2/ESCO1, and STAG2/CTCF-depleted HeLa cells. Loops were identified by HiCCUPS.
D. Aggregate TAD analysis for the conditions described in (C). Average coverage-corrected Hi-C contact matrices are shown around 318x 390-410 kb long TADs identified in the control-depleted cells.
E. Coverage-corrected Hi-C contact matrices of the whole of chromosome 2 from cells treated as in (C). Matrices were plotted using Juicebox.
F. Long range (>2 Mb) intra-chromosomal (cis) and inter-chromosomal (trans) contact enrichment between bins with varying compartment signal strength from most B-like (1) to most A-like (50) from HeLa cells depleted for proteins as in (A).

Table 1. Generation of cell lines used in this study

cell line	original cell line	gene	tag	method	reference or gRNAs
SMC3-LAP	HeLa	SMC3	LAP tag at the C-terminus	BAC rcombiniering	Poser et al., 2008
EGFP-STAG1 cl H4	HeLa	STAG1	EGFP at the N-terminus	CRISPR-Cas9	ACAATACTTACTGTAACACtgg and TATTTTTTAAGGAAAATTTtgg
STAG1-EGFP H8	HeLa	STAG1	EGFP at the C-terminus	CRISPR-Cas9	CACCGACAATACTTACTGTAACAC and AAACAAATTTTCCTTAAAAAATAC
STAG2-EGFP F2	HeLa	STAG2	EGFP at the C-terminus	CRISPR-Cas9	CACCGCACAGATTTAATTGTGTAC and AAACCAGATTTAATTGTGTACTGC
EGFP-STAG1 FLAG-RFP-STAG2 clD10	HeLa	STAG1,STAG2	EGFP at the N-terminus of STAG1;	CRISPR-Cas9	ACAATACTTACTGTAACACtgg and TATTTTTTAAGGAAAATTTtgg
			FLAG-RFP at the N-terminus of STAG2	CRISPR-Cas9	ATTTACGTGGGTAAAATGGtgg and GAATATATTTCTGACATTGagg
SMC1-LAP	MEFs	SMC1	LAP tag at the C-terminus	BAC rcombiniering	Poser et al., 2008
SCC1-LAP	MEFs	SCC1	LAP tag at the C-terminus	BAC rcombiniering	Poser et al., 2008

Table 2. Summary statistics for Hi-C data sets generated in this study

No.	Condition	Biological replicate	Restriction enzyme	Sample name	Total read pairs[a]	Unique read pairs[b]	trans read pairs[c]	% trans read pairs[c]	*Cis* comp.score[d]	*Trans* comp.score[e]	TAD genome coverage[f] (%)	number of TADs[g]	median TAD length (kb)[h]	Average TADB strength (insulation score)[i1]	Average TADB strength (insulation score)[i2]	number of loops[j]
1	control high resolution	1	MboI	HeLa_control__1_2	1.474.037.625	916.286.586	82.912.240	9,0%	0,44	0,33	71,8	4.092	400.000	1,03	1,03	24.444
2	STAG1 RNAi high resolution	1	MboI	SA1_siRNA__1_2	1.369.342.451	855.693.522	119.697.796	14,0%	0,38	0,29	72,0	3.702	440.000	0,95	1,05	11.545
3	STAG2 RNAi high resolution	1	MboI	SA2_siRNA__1_2	1.334.974.287	840.729.828	115.715.018	13,8%	0,37	0,27	73,5	3.440	490.000	0,84	1,03	14.047
4	control R1	1	HindIII	HeLa_G1_control_r1	199.688.663	98.211.144	5.425.643	5,5%	0,56	0,48	58,9	3.144	445.000	0,91	0,91	10844
		2	HindIII	HeLa_G1_control_r2	240.143.927	108.271.389	5.783.217	5,3%	0,51	0,47	58,2	3.089	450.000	0,90	0,89	
5	CTCF RNAi	1	HindIII	HeLa_G1_CTCF_r1	237.323.139	108.135.338	9.664.980	8,9%	0,31	0,19	62,4	2.656	567.500	0,63	0,87	4414
		2	HindIII	HeLa_G1_CTCF_r2	291.987.724	133.761.550	13.171.488	9,8%	0,32	0,22	63,3	2.670	585.000	0,60	0,86	
6	ESCO1 RNAi	1	HindIII	HeLa_G1_ESCO1_r1	215.477.899	94.411.272	13.854.484	14,7%	0,50	0,43	63,8	2.760	560.000	0,76	0,96	3555
		2	HindIII	HeLa_G1_ESCO1_r2	234.853.785	85.463.840	12.859.640	15,0%	0,51	0,41	63,2	2.730	565.000	0,75	0,96	
7	STAG1 siRNA	1	HindIII	HeLa_G1_SA1_r1	189.078.810	95.572.782	9.453.126	9,9%	0,49	0,40	60,9	2.832	530.000	0,78	0,89	6834
		2	HindIII	HeLa_G1_SA1_r2	288.779.039	121.694.728	10.026.275	8,2%	0,45	0,38	59,4	2.767	540.000	0,79	0,87	
8	STAG2 siRNA	1	HindIII	HeLa_G1_SA2_r1	225.983.542	111.684.868	9.470.061	8,5%	0,47	0,37	62,1	2.704	565.000	0,76	0,95	11378
		2	HindIII	HeLa_G1_SA2_r2	272.546.577	136.916.058	10.711.215	7,8%	0,43	0,35	61,8	2.671	570.000	0,75	0,91	
9	STAG1+CTCF	1	HindIII	HeLa_G1_CTCF_SA1_r1	293.873.827	137.409.357	15.064.145	11,0%	0,36	0,28	63,3	2.649	585.000	0,55	0,83	2621
		2	HindIII	HeLa_G1_CTCF_SA1_r2	230.498.643	99.063.364	10.530.598	10,6%	0,38	0,26	63,0	2.666	575.000	0,58	0,85	
10	STAG2+CTCF	1	HindIII	HeLa_G1_CTCF_SA2_r1	202.518.989	91.072.235	11.213.070	12,3%	0,32	0,13	63,3	2.486	630.000	0,57	0,89	3439
		2	HindIII	HeLa_G1_CTCF_SA2_r2	276.301.529	111.224.693	14.928.679	13,4%	0,31	0,10	62,1	2.388	655.000	0,57	0,87	
11	ESCO1 + SA2 RNAi	1	HindIII	HeLa_G1_ESCO1_SA2_r1	373.623.777	217.034.789	36.025.922	16,6%	0,42	0,38	66,1	2.555	654.375	0,61	0,86	6834
		2	HindIII	HeLa_G1_ESCO1_SA2_r2	248.368.126	126.629.139	19.871.331	15,7%	0,46	0,40	63,2	2.710	570.000	0,67	0,92	

a: raw number of read pairs from paired-end sequencing
b: unique valid mapped read pairs from HiCUP v0.5.8
c: % of unique valid read pairs that are inter-chromosomal
d: log2 contact enrichment of A-A and B-B contacts for long-range (>10Mb) intra-chromosomal contacts
e: log2 contact enrichment of A-A and B-B contacts for inter-chromosomal contacts
f: % of genome covered by TADs called by HOMER v4.7
g: number of TADs called by HOMER v4.7
h: number of loops called by the HiCCUPS algorithm of Juicer tools v0.7.5
i1: average standardised insulation score at the corresponding G1 control TAD boundaries (hires or r1,r2 average) called by HOMER v4.7 in the respective conditions
i2: average standardised insulation score at the TAD boundaries called by HOMER v4.7 in the respective conditions
j: number of loops called by the HiCCUPS algorithm of Juicer tools v0.7.5; please note that the number of loops that can be called depends on the number of unique read pairsthis needs to be taken into consideration when comparing corner peaks between ddifferent experiments

Table 3: Number of cells analysed by FISH and statistical significance

Domain 1						
condition	**control**	**CTCF**	**SCC1**	**STAG1**	**STAG2**	**STAG1/2**
n	648	382	681	561	533	516
t-test to control	1	<0.0001	<0.0001	<0.0001	0.8317	<0.0001

Domain 2						
condition	**control**	**CTCF**	**SCC1**	**STAG1**	**STAG2**	**STAG1/2**
n	267	259	154	179	215	118
t-test to control	1	<0.0001	<0.0001	<0.0001	0.2478	<0.0001

Domain 3						
condition	**control**	**CTCF**	**SCC1**	**STAG1**	**STAG2**	**STAG1/2**
n	432	382	842	698	605	603
t-test to control	1	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001

Domain 4						
condition	**control**	**CTCF**	**SCC1**	**STAG1**	**STAG2**	**STAG1/2**
n	108	127	141	131	173	176
t-test to control	1	<0.0001	<0.0001	<0.0001	0.5345	<0.0001

Domain 5						
condition	**control**	**CTCF**	**SCC1**	**STAG1**	**STAG2**	**STAG1/2**
n	380	366	165	481	280	565
t-test to control	1	0.6979	<0.0001	<0.0001	<0.0001	<0.0001

Domain 6						
condition	**control**	**CTCF**	**SCC1**	**STAG1**	**STAG2**	**STAG1/2**
n	201	295	368	311	256	387
t-test to control	1	<0.0001	<0.0001	<0.0001	0.0941	<0.0001

Negative control						
condition	**control**	**CTCF**	**SCC1**	**STAG1**	**STAG2**	**STAG1/2**
n	354	124	124	142	130	139
t-test to control	1	0.5165	0.6555	0.6759	<0.0001	0.1996