

## **Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF**

Abrar Aljahani<sup>1</sup>, Peng Hua<sup>2</sup>, Magdalena A. Karpinska<sup>1</sup>, Kimberly Quililan<sup>1</sup>, James O.J. Davies<sup>2,\*</sup>, A. Marieke Oudelaar<sup>1,\*</sup>

1. Max Planck Institute for Biophysical Chemistry, Göttingen, Germany

2. MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK

\* Corresponding authors: [james.davies@imm.ox.ac.uk](mailto:james.davies@imm.ox.ac.uk), [marieke.oudelaar@mpibpc.mpg.de](mailto:marieke.oudelaar@mpibpc.mpg.de)

## ABSTRACT

Enhancers and promoters predominantly interact within large-scale topologically associating domains (TADs), which are formed by loop extrusion mediated by cohesin and CTCF. However, it is unclear whether complex chromatin structures exist at sub-kilobase-scale and to what extent fine-scale regulatory interactions depend on loop extrusion. To address these questions, we present an MNase-based chromosome conformation capture (3C) approach, which has enabled us to generate the most detailed local interaction data to date and precisely investigate the effects of cohesin and CTCF depletion on chromatin architecture. Our data reveal that *cis*-regulatory elements have distinct internal nano-scale structures, within which local insulation is dependent on CTCF, but which are independent of cohesin. In contrast, we find that depletion of cohesin causes a subtle reduction in longer-range enhancer-promoter interactions and that CTCF depletion can cause rewiring of regulatory contacts. Together, our data show that loop extrusion is not essential for enhancer-promoter interactions, but contributes to their robustness and specificity and to precise regulation of gene expression.

## MAIN

Mammalian gene expression patterns are controlled by enhancers, which form specific interactions with the promoters of their target genes to transfer activating signals<sup>1</sup>. Since these elements can be separated by large genomic distances, the specificity of enhancer-promoter interactions is dependent on the 3D organization of the genome<sup>2</sup>. Loop extrusion, mediated by cohesin and CTCF, organizes the genome into TADs<sup>3-6</sup>. Regulatory interactions between enhancers and promoters predominantly occur within TADs<sup>7</sup> and disruption of TAD boundaries can result in ectopic enhancer-promoter interactions and changes in gene expression<sup>3,8-16</sup>.

It has been shown that depletion of cohesin, its loading factor NIPBL, and CTCF result in large-scale changes in the 3D organization of the genome and a loss of TADs<sup>17-23</sup>. However, surprisingly, the effects of these depletions on gene expression have been reported to be relatively mild. This poses an important question: to what extent are the specific regulatory interactions between enhancers and promoters that mediate gene expression dependent on loop extrusion?

To date, the effects of depletion of cohesin and CTCF on 3D chromatin structure have been analyzed by genome-wide Chromosome Conformation Capture (3C)<sup>17-22,24</sup> and imaging approaches<sup>21,25-27</sup>. Although these studies have given important insights into the roles of cohesin and CTCF in organizing the genome, the resolution and sensitivity of the approaches used in these studies is not sufficient to characterize potential changes in sub-kilobase-scale chromatin architecture and enhancer-promoter interactions upon cohesin and CTCF depletion.

Conventional 3C methods<sup>28,29</sup> are limited in their resolution by sequencing depth, library complexity and the distribution of the recognition sites of the restriction enzymes used for chromatin digestion<sup>30</sup>. The latter can be overcome by the use of MNase, which digests the genome largely independent of DNA sequence. MNase digestion was initially implemented in the Micro-C approach in yeast<sup>31,32</sup>. Micro-C has recently been adapted for use in mammalian genomes and has demonstrated a better signal-to-noise ratio compared to Hi-C and potential to generate contact matrices at high resolution<sup>33-35</sup>.

We have recently developed a novel MNase-based approach called Micro-Capture-C (MCC), which is capable of generating base pair resolution 3C data in mammalian cells<sup>36</sup>. This technique produces interaction profiles for selected viewpoints, but does not give an overview of the complete 3D structure of regions of interest. We have therefore extended the MCC approach and combined it with enrichment panels consisting of capture oligonucleotides tiled across genomic regions of interest (Figure 1a). This Tiled-MCC approach has a number of advantages over Micro-C, which enable the generation of data at higher resolution. These include maintaining cellular and nuclear architecture by using digitonin instead of traditional detergents; very deep data generation through high library complexity and targeted sequencing; and direct identification of ligation junctions. The Tiled-MCC approach has enabled us to generate the deepest local 3C matrices to date in mouse embryonic stem (mES) cells with high reproducibility (Supplementary Figure 1). Direct comparison of Tiled-MCC and Micro-C<sup>34</sup> data clearly shows that Tiled-MCC is able to detect enhancer-promoter interactions and long-range contacts between CTCF-binding elements which are not detected by Micro-C (Figure 1b-d).

Since Tiled-MCC libraries are sonicated to an average size of ~200 base pairs and sequenced with paired-end reads of 150 base pairs each, ligation junctions are sequenced directly and can be precisely reconstructed bioinformatically. This further increases the signal-to-noise ratio of the Tiled-MCC contact matrices and allows for generation of density plots of ligation junctions in which local chromatin

structures can be analyzed with unprecedented resolution. These density plots reveal characteristic micro-topologies at *cis*-regulatory elements (Figure 1e, Supplementary Figure 2). Promoters and super-enhancers form complex nano-scale structures, characterized by sub-kilobase compartments of interaction. CTCF-binding elements form distinct patterns of phased nucleosomes and localized stripes. Interestingly, CTCF-binding sites within super-enhancers create localized regions of insulation between the different micro-compartments. Non-active chromatin is characterized by regular nucleosome structures without specific phasing patterns.

The advantages of Tiled-MCC make it a unique approach to investigate changes in local chromatin structure in detail upon depletion of architectural proteins. To study the role of cohesin and CTCF in mediating large- and fine-scale chromatin structure, we used mES cell lines in which the cohesin subunit RAD21 (SCC1) or CTCF can be rapidly degraded via an auxin-inducible degron (AID). To examine enhancer-promoter interactions in the absence of cohesin, we produced Tiled-MCC matrices of four well-characterized gene loci in RAD21-AID mES cells. We initially focused our analyses on the region containing the *Sox2* gene and its super-enhancer<sup>37</sup>. Consistent with previous reports<sup>18,20</sup>, cohesin depletion results in a near-complete loss of TAD structure and long-range interactions between CTCF-binding sites (Figure 2a, Supplementary Figure 3). A close-up view reveals that interactions between the *Sox2* promoter and its super-enhancer remain partially intact, despite loss of the interactions between the CTCF-binding sites surrounding these elements (Figure 2b). However, the interactions between the enhancers and promoter are reduced in intensity. This reduction in enhancer-promoter contacts corresponds to a small, but significant decrease in *Sox2* expression (Figure 2c). We next investigated the *Prdm14*, *Nanog* and *Pou5f1* loci, which are also regulated by well-characterized super-enhancers. In each of these loci we find a subtle decrease in enhancer-promoter interactions and gene expression following cohesin depletion (Supplementary Figure 4).

Together, these results indicate that cohesin contributes to enhancer-promoter interactions. However, it is not clear if the effects of cohesin depletion are due to loss of TAD insulation or result from loss of active loop extrusion. To get further insight into the underlying mechanism, we studied the effects of CTCF depletion, which leads to a loss of TAD boundaries, but does not interfere with loop extrusion otherwise. Consistent with the literature<sup>19,20</sup>, Tiled-MCC data in CTCF-AID mES cells show that CTCF depletion results in a loss of TADs and interactions between CTCF-binding sites (Figure 3, Supplementary Figure 5). Furthermore, at the *Nanog* locus, we find that loss of CTCF leads to rewiring of the interactions between enhancers and promoters (Figure 3). Upon CTCF depletion, the enhancers form stronger contacts with the *cis*-regulatory elements of the downstream *Foxj2* gene, which leads to upregulation of its expression. We also observe the formation of ectopic enhancer-promoter interactions at the *Prdm14* locus (Supplementary Figure 6), which are consistent with previously reported effects of CTCF-binding site deletions in this locus<sup>8,38</sup>. However, CTCF depletion at the *Sox2* and *Oct4* loci does not result in changes in enhancer-promoter interactions or gene expression in mES cells (Supplementary Figure 6). These results show that CTCF is important for the specificity of enhancer-promoter interactions, but not required for their formation or maintenance.

To examine the impact of cohesin and CTCF depletion on *cis*-regulatory elements in further detail, we analyzed the effects on local micro-topologies (Figure 4, Supplementary Figure 7). As expected, the pattern of phased nucleosomes, stripes and insulation at CTCF-binding elements are lost upon CTCF depletion. However, surprisingly, the localized insulating properties are not altered by depletion of cohesin, despite complete loss of long-range contacts between CTCF-binding sites upon cohesin depletion (Figure 4, Supplementary Figures 7 and 8). The nano-scale structures of promoters and super-enhancers are also not affected by cohesin depletion, but insulation at CTCF-binding sites within super-



enhancers or promoter regions is lost upon CTCF depletion (Figure 4, Supplementary Figure 7). This suggests that CTCF mediates local insulation independent of cohesin-mediated loop extrusion, possibly via recruitment of other factors or its effects on the organization of nucleosomes.

These observations are confirmed by micro-topology analysis of a larger region at the *Nanog* locus, which contains a promoter, an enhancer and both a strong and a weak CTCF-binding site (Figure 5). Upon cohesin depletion, the specific interactions between the CTCF-binding sites are lost, but the stripe pattern at the strong CTCF-binding site remains, whereas CTCF depletion results in loss of both features. In contrast, the local structures at the enhancer and promoter are not dependent on cohesin and CTCF.

Together, our data show that *cis*-regulatory elements are characterized by distinct internal nano-scale structures, which are independent of cohesin. Interestingly, this includes localized insulation mediated by CTCF-binding sites. However, long-range contacts between CTCF-binding elements and TAD insulation are largely abrogated by cohesin depletion. In contrast, interactions between enhancer and promoters are still present in the absence of cohesin. This is consistent with the findings of a recent preprint<sup>39</sup>. However, our data show that enhancer-promoter interactions are reduced upon cohesin depletion, which indicates that cohesin does contribute to the robustness of enhancer-promoter interactions. This could be mediated by an increase in their interaction frequency through the process of loop extrusion or result from direct bridging of enhancers and promoters by cohesin. These findings are consistent with a recently reported role of the cohesin release factor WAPL in regulating gene expression<sup>40</sup> and a potential role of cohesin movement along the chromatin in directing regulatory factors, as suggested in a recent preprint<sup>41</sup>. Of interest, it has been suggested that cohesin might be particularly important for mediating very long-range interactions between enhancers and promoters<sup>24,42</sup>. While our data show little change in enhancer-promoter interactions across very small distances (<5 kb), we see a decrease in enhancer-promoter interactions at the relatively small *Pou5f1* and *Nanog* loci, across distances of ~20 and ~50kb, respectively. Furthermore, consistent with the literature, our data show that the CTCF boundaries of TADs contribute to the specificity of enhancer-promoter interactions. Importantly, both the rewiring of enhancer-promoter interactions upon CTCF depletion and the reduced enhancer-promoter interactions upon cohesin depletion are associated with changes in gene expression.

In our study, we have focused on gene loci that are specifically expressed in mES cells. We find consistent reductions in enhancer-promoter interactions and gene expression upon cohesin depletion in all investigated loci. This appears to be in contrast with reported mild effects on gene expression upon loss of cohesin or NIPBL<sup>17,18</sup>. However, our data suggest that cohesin is particularly important for mediating robustness of enhancer-promoter interactions. It is therefore perhaps not surprising that only a subset of genes is mis-regulated upon cohesin depletion, since loss of cohesin might only effect genes whose expression is critically dependent on the robust formation of tissue-specific enhancer-promoter interactions. This is consistent with the reported importance of cohesin for regulating gene expression upon stimuli and during differentiation<sup>43,44</sup>.

Taken together, our high-resolution data identify the existence of complex nano-scale structures within *cis*-regulatory elements, which are independent of loop extrusion. Furthermore, we show that loop extrusion is not essential for the formation of longer-range enhancer-promoter interactions, but an important mechanism to regulate the robustness and specificity of regulatory interactions.

## METHODS

### *Cell culture*

Wild type (E14), RAD21-mAID-eGFP<sup>21</sup>, and CTCF-AID-eGFP<sup>19</sup> mouse embryonic stem (mES) cells were cultured on plates pre-coated with 0.1% gelatin (Sigma-Aldrich, G1393) in Glasgow Modified Essential Medium (Gibco, 11710035) supplemented with 10% Fetal Bovine Serum (Gibco, 10270106), 0.01 mM 2-Mercaptoethanol (Gibco, 31350010), 2.4 mM L-glutamine (Gibco, 25030024), 1X non-essential amino acids (Gibco, 11140050), 1X sodium pyruvate (Gibco, 11360070) and 20 ng/ml recombinant mouse Leukemia-Inhibitory Factor (Cell Guidance Systems, GFM200). Cells were grown at 37 °C and 5% CO<sub>2</sub>. Half of the media was exchanged daily and cells were passaged every 2 days by trypsinization (Gibco, 25300054).

### *Depletion of RAD21 and CTCF*

RAD21-mAID-eGFP and CTCF-AID-eGFP mES cells were passaged and plated 2 days before auxin treatment. On the treatment day, cells were first washed once with phosphate-buffered saline (PBS) and supplemented with equilibrated (37 °C and 5% CO<sub>2</sub>) medium containing 500 µM indole-3-acetic acid (Sigma-Aldrich, I5148) freshly prepared before use. RAD21-mAID-eGFP cells were treated with auxin for 6 hr. Since complete depletion of CTCF takes longer to achieve<sup>19,45</sup>, CTCF-AID-eGFP cells were treated for 48 hr, with exchanging auxin-containing media after the first 24 hr. Alkaline phosphatase testing for the assessment of pluripotency before and after auxin treatment was performed using the Alkaline Phosphatase Detection Kit (EMD Millipore, SCR004).

### *Immunoblotting*

RAD21-mAID-eGFP and CTCF-AID-eGFP mES cells were grown in 25 cm<sup>2</sup> flasks until they reached ~80% confluency. Cells were then treated with auxin following the aforementioned treatment conditions for each cell line. Whole-cell extracts were prepared by first dissociating cells by trypsinization, resuspending them in media, pelleting and washing once with PBS. The pellet was then resuspended in 5X the pellet volume in radioimmunoprecipitation assay buffer (RIPA; Thermo Scientific, 89900) with protease inhibitor cocktail (Leupeptin, Carl Roth, CN33.4; Pepstatin A, Carl Roth, 2936.3; PMSF, Carl Roth, 6367.3; Benzamidine hydrochloride, Acros Organics, 105245000, resuspended in ethanol) and 250 U/µL benzonase (Sigma-Aldrich, E1014) and rotated for 1 hr at 4 °C. After centrifugation at maximum speed, whole-cell extract in the supernatant was measured using the Bio-Rad Protein Assay (Bio-Rad, 5000006). 15 µg of protein extracts were mixed with NuPAGE LDS Sample Buffer (Invitrogen, NP0007), supplemented with 50 mM DTT (Carl Roth, 6908.3), and loaded on a NuPAGE 4%-12% Bis-Tris gel (Invitrogen, NP0321PK2). Proteins were transferred to a PVDF membrane using the XCell Blot Module (Invitrogen, EI9051) for 1 hr at 30 V. Membrane blocking was performed in 5% blotting grade milk powder (Carl Roth, T145.2) dissolved in PBS-0.05% Tween-20 (PBS-T) for 1 hr at room temperature. After rinsing the membranes with PBS-T, incubation with primary antibodies was performed as follows: the membrane was cut into two pieces, one piece corresponding to the higher protein ladder size for the incubation of RAD21 and CTCF primary antibodies (RAD21: anti-RAD21 antibody, rabbit polyclonal, Abcam, ab154769, 1:1000; CTCF: anti-CTCF antibody, rabbit polyclonal, Abcam, ab70303, 1:1000), and the second piece with lower protein ladder size for the incubation with the loading control (Histone H3: anti-histone H3 antibody (HRP), Abcam, ab21054, 1:5000). All primary antibodies were diluted in 2% PBS-T and incubated at 4 °C overnight. The next morning, membranes were washed 3X with PBS-T for 5-10 minutes each and incubated with secondary antibody (RAD21 and CTCF: Goat Anti-Rabbit IgG H&L (HRP), Abcam

ab205718) in PBS-T in a 1:5000 dilution for 1 hr at room temperature, washed again 3X and analyzed on the Intas ChemoCam Imager HR.

### ***Tiled-MCC – Fixation***

The preparation of Micro-Capture-C (MCC) libraries was performed as previously described<sup>36</sup>. 10 million cells per biological replicate were washed once with PBS, dissociated with trypsin and resuspended in 10 ml of culture media. Fixation was performed by incubating cells with formaldehyde to a final concentration of 2% on a roller mixer for 10 minutes at room temperature. The reaction was quenched by adding cold glycine to a final concentration of 130 mM. Samples were centrifuged at 200 rcf for 5 minutes at 4 °C, the pellet was washed once with cold PBS, centrifuged again and resuspended in 1 ml of cold PBS containing 0.005% digitonin (Sigma-Aldrich, D141).

### ***Tiled-MCC – Library preparation***

Crosslinked and permeabilized cells were centrifuged at 300 rcf for 5 minutes. The supernatant was carefully removed without disrupting the pellet, which was subsequently resuspended in 900 µl nuclease-free water. Cells were then split equally into 3 digestion reactions, such that each reaction contained 3-4 million cells. Titration of different MNase concentrations (NEB, M0247) was performed for each aliquot with MNase concentrations ranging from 30–60 Kunitz U in a total reaction volume of 800 µl containing low-calcium MNase buffer (50 mM Tris-HCl pH 7.5, 1 mM CaCl<sub>2</sub>). The reaction was then incubated for 1 hr at 37 °C on an Eppendorf Thermomixer at 550 rpm, after which it was quenched with 5 mM of ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid (EGTA, Sigma, E3889). The quenched reaction was subsequently centrifuged for 5 minutes at 300 rcf and the supernatant was carefully discarded. The pellet was resuspended in 1 ml of PBS with 5 mM EGTA. 200 µl was removed as a control for MNase digestion, from which DNA was extracted using the DNeasy blood and tissue kit (Qiagen, 69504) and digestion efficiency was assessed using the Agilent D1000 TapeStation (Agilent Technologies, 5067-5582). MNase reactions that yielded ~180 bp fragments, corresponding to mono-nucleosomes with linkers, were taken further for the subsequent reactions.

To minimize DNA loss, DNA end repair, phosphorylation and ligation were performed in a single tube. However, by controlling the temperature at which the respective enzymes are active, end repair and phosphorylation were performed before ligation. The remaining 800 µl of MNase-digested chromatin resuspended in PBS and EGTA was centrifuged for 5 minutes at 300 rcf. The supernatant was carefully discarded and the pellet was resuspended in DNA ligase buffer (Thermo Scientific, B69) supplemented with dNTPs (NEB, N0447L) at a final concentration of 400 µM each and 2.5 mM EGTA. T4 Polynucleotide Kinase (NEB, M0201L), DNA Polymerase I Large Fragment (Klenow, NEB, M0210L), and T4 DNA ligase (Thermo Scientific, EL0013) were added to final concentrations of 200 U/ml, 100 U/ml, and 300 U/ml, respectively. The reaction was incubated on an Eppendorf Thermomixer at 550 rpm for 2 hr at 37 °C, followed by a 16-hour incubation at 20 °C. Decrosslinking of chromatin was performed using proteinase K (included in the Qiagen DNeasy blood and tissue kit) at 65 °C for > 4 hr, which was followed by DNA extraction using the Qiagen DNeasy blood and tissue kit. The ligation product, referred to as 3C library hereafter, was assessed using the Agilent D1000 TapeStation. A successful ligation is indicated by a significant increase in the fragment size > ~370 bp.

### ***Tiled-MCC – Sonication and ligation of indexed sequencing adapters***

Sonication of 3C libraries was performed using 3-5 µg per library on a Covaris S220 Focused Ultrasonicator with the following conditions: 250-300s: duty cycle 10%; intensity 5; cycles per burst 200, to yield an average fragment size of 200 bp. The sonication quality was assessed using the Agilent D1000 TapeStation. The DNA was purified using Ampure XP beads (Beckman Coulter, A63881). To

maximize library complexity, the addition of sequencing adapters was parallelized in triplicate reactions such that each reaction contained 1-2 µg of sonicated 3C library. NEB Ultra II (NEB, 7645S) reagents were used following the manufacturer's protocol with the following deviations: (1) 2-3 X the amount of adapters was used; (2) all Ampure XP bead clean-up reactions were performed with a DNA sample:bead ratio of 1:1.5; (3) to maximize library complexity and yield, the PCR was performed in triplicate per ligation reaction using the Herculase II PCR reagents (Agilent Technologies, 600677). The parallel library preparations and PCR reactions were subsequently pooled for each reaction.

### ***Tiled-MCC – Capture oligonucleotide design***

Tiled-MCC capture panels were designed to densely cover regions of interest, with oligonucleotides that are 70 nucleotides in length and have an overlap of 35 nucleotides. The sequences were designed and filtered for repetitive sequences using a python-based oligo tool<sup>46</sup> (<https://oligo.readthedocs.io/en/latest/>). The panels of double-stranded capture oligonucleotides were ordered from Twist Bioscience (Custom probes for NGS target enrichment).

### ***Tiled-MCC – Enrichment***

The enrichment procedure was performed using the Twist Hybridization and Wash Kit (Twist Bioscience, 101025), Twist Universal Blockers (Twist Bioscience, 100578), and Twist Binding and Purification Beads (Twist Bioscience, 100983). Per hybridization reaction, up to 8 amplified and indexed libraries were multiplexed to a final amount of 1.5 µg in a single 0.2-ml PCR strip-tube. Library pools were dried completely in a vacuum concentrator at 45 °C. Dried DNA was resuspended in 5 µg of mouse Cot-1 DNA and 7 µl of Twist Universal Blockers. In a separate PCR 0.2-ml strip-tube, the probe solution was prepared by mixing 20 µl of Twist Hybridization Mix with 1-2 µl of oligonucleotides and the final volume was adjusted with nuclease-free water. To prepare both the probe solution and the resuspended indexed library pool for hybridization, the probe solution was heated to 95 °C for 2 minutes in a PCR thermal cycler with the lid at 105 °C, then immediately cooled down on ice for 5 minutes. While the probe solution was cooling down on ice, the library pool was heated following the same conditions for 5 minutes. After equilibrating both mixtures to room temperature for 1-2 minutes, the probe solution was added to the library pool, and 30 µl of Twist Hybridization Enhancer was added last. The capture reaction was incubated at 70 °C for 16 hr in a PCR thermal cycler with the lid heated to 85 °C. The hybridization reaction was subsequently mixed with Streptavidin Binding Beads for 30 minutes at room temperature on a shaker. Washing with Twist Wash Buffers 1, 2 and 3 was performed following the Twist Target Enrichment Protocol. Post-hybridization PCR was performed with 11-12 amplification cycles. Enriched library was purified with pre-equilibrated Twist DNA Purification Beads at a ratio of 1:1.8 DNA to beads. DNA quantification and QC validation were performed using the Qubit dsDNA Broad Range Quantification Assay (Life Technologies, Q32850) and the Agilent Bioanalyzer Broad Range DNA kit (Agilent Technologies, 5067-5582), respectively.

### ***Tiled-MCC – Sequencing***

Libraries were sequenced using the Illumina NovaSeq and NextSeq 550 platforms with 150 bp paired-end reads. Depending on the quality of the MCC libraries, sequencing 40-200 million reads per enriched Mb per pooled sample is sufficient for data at 500 bp resolution.

### ***Tiled-MCC – Analysis***

Tiled-MCC analysis was performed using the MCC pipeline<sup>36</sup> (<https://github.com/jojdavies/Micro-Capture-C>). Briefly, adapter sequences were removed using Trim Galore (Babraham Institute, v.0.3.1) and paired-end reads were reconstructed using FLASH<sup>47</sup> (v.1.2.11). Reads were then mapped to the DNA sequences corresponding to the enriched regions of interest with the non-stringent aligner BLAT<sup>48</sup>

(v.35), using a custom-made file containing the sequences of the regions of interest as reference. Based on the mapping by BLAT, the reads in the FASTQ files were split into two or more reads corresponding to the chimeric fragments, and into different files based on the region they mapped to, using the MCCsplitter.pl script. Uninformative reads that did not contain a sequence that mapped to any of the enriched regions of interest were discarded. The split reads in the FASTQ files were subsequently mapped to the mm10 reference genome using Bowtie2<sup>49</sup> (v.2.3.5). The aligned reads were further processed using the MCCanalyser.pl script. PCR duplicates were removed based on the sonicated ends and ligation junction and allowing for a wobble of  $\pm 2$  bp. Unique ligation junctions were identified if the fragment ends were less than 5 bp apart in the original read and separated by mapping with BLAT and Bowtie2.

To generate contact matrices, the unique ligation junctions were converted into raw matrices, which were balanced using ICE normalization<sup>50</sup>. To generate density plots, the unique ligation junctions were filtered for a minimum distance  $\geq 10$  bp and plotted in python as a scatter plot with a color code defined by the local density of the data points.

The data presented in the manuscript represent the averages of 9 replicates for WT samples, 4 replicates for untreated RAD21-AID samples, 6 replicates for auxin-treated RAD21-AID samples, and 6 replicates for auxin-treated CTCF-AID samples. For direct comparisons between Tiled-MCC matrices or density plots, the data were down-sampled to the lowest number of filtered, unique ligation junctions per condition or replicate. The RAD21-AID and CTCF-AID cell lines are derived from the same strain as the WT cells and contact matrices from the untreated RAD21-AID samples are nearly identical to the WT samples (Supplementary Figure 3b). The contact matrices presented in the manuscript therefore show comparisons between auxin-treated RAD21-AID and CTCF-AID samples and WT samples.

### ***Micro-C analysis***

The Micro-C data were downloaded from GSE130275<sup>34</sup> as files listing valid pairs in the mm10 reference genome, which were converted into ICE-normalized<sup>50</sup> contact matrices using HiC-Pro<sup>51</sup>.

### ***ChIP-seq analysis***

Alignment to the mm10 reference genome and processing of ChIP-seq data was performed using the NGseqBasic pipeline<sup>52</sup>. ChIP-seq data for CTCF, RAD21, and H3K4me1 were downloaded from GSE30203<sup>53</sup>, GSE94452<sup>22</sup> and GSE27844<sup>54</sup>, respectively. ChIP-seq data for H3K27ac and H3K4me3 and DNase I hypersensitivity data for mES cells were accessed via ENCODE<sup>55</sup>.

### ***RNA-seq analysis***

The normalized read counts, P-values, and significance scores for genes of interest in RAD21-AID<sup>21</sup> and CTCF-AID<sup>19</sup> mES cells were extracted from the original processed data files shared by the authors of the respective articles.

## **ACKNOWLEDGEMENTS**

A.M.O. is funded by the Max Planck Society. J.O.J.D. is funded by an MRC Clinician Scientist Award (MRC Clinician Scientist Fellowship MR/R008108). We would like to thank Angelika Feldmann, James Rhodes and Rob Klose for providing the RAD21-AID mES cell line and Elphège Nora and Benoit Bruneau for providing the CTCF-AID mES cell line. We are grateful to Patrick Cramer for advice and infrastructure support, Michael Lidschreiber for assistance with bioinformatic analysis, Taras Velychko for support with experiments, and Kerstin Maier and Petra Rus for help with sequencing. We would like to thank Douglas Higgs, Jim Hughes, Damien Downes and Elisa Oberbeckmann for advice and feedback on the manuscript.

## **COMPETING INTERESTS**

J.O.J.D. is a co-founder of Nucleome Therapeutics and provides consultancy to the company.

## **DATA AND CODE AVAILABILITY**

Sequencing data have been submitted to the NCBI Gene Expression Omnibus under accession code GSE181694. Scripts for Tiled-MCC analysis are available for academic use through the Oxford University Innovation software store (<https://process.innovation.ox.ac.uk/software/p/16529a/micro-capture-c-academic/1>).



## REFERENCES

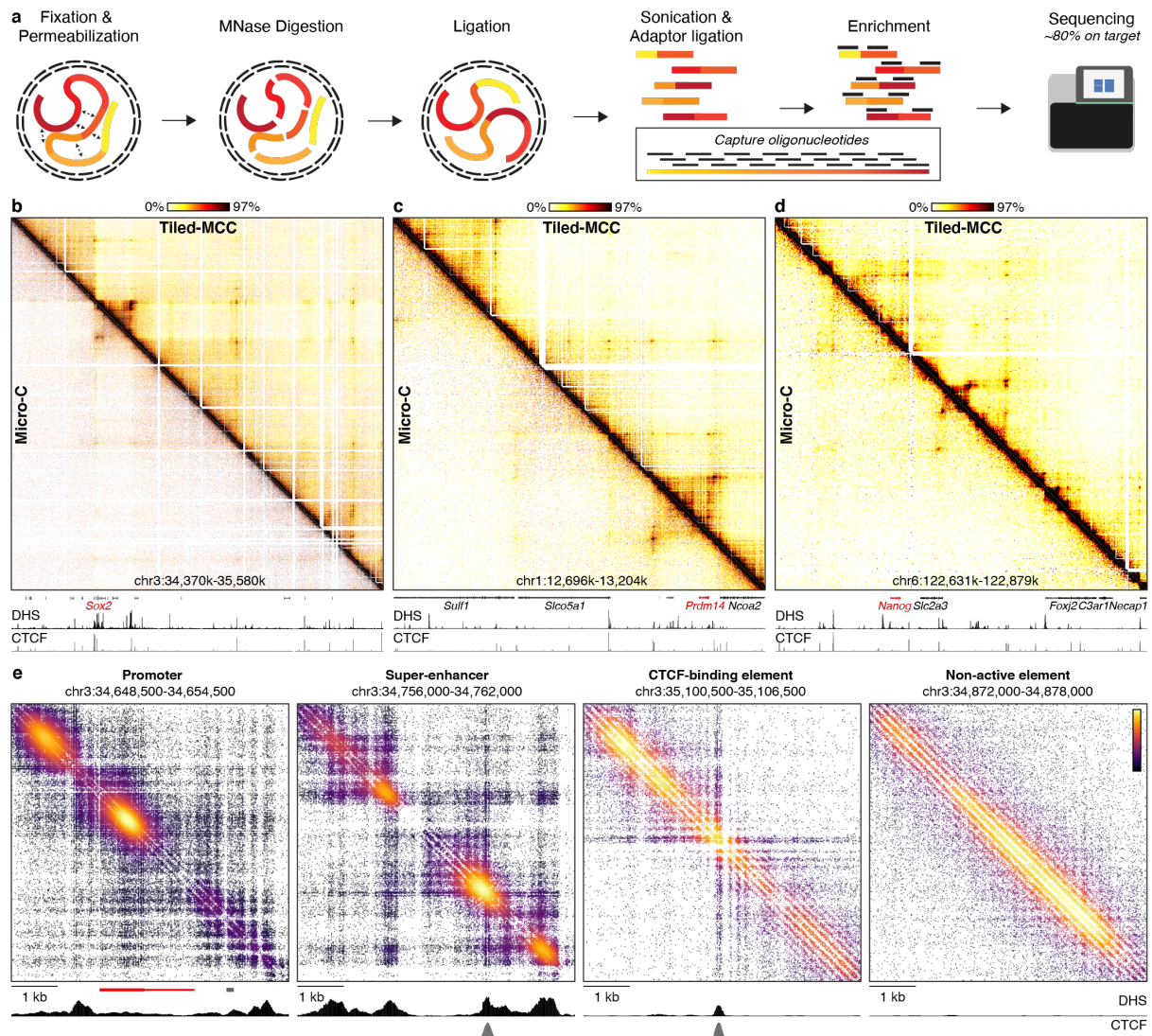
- 1 Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187, doi:<https://doi.org/10.1016/j.cell.2016.09.018> (2016).
- 2 Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. *Nature Reviews Genetics* **22**, 154-168, doi:10.1038/s41576-020-00303-x (2021).
- 3 Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6456-6465, doi:10.1073/pnas.1518552112 (2015).
- 4 Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *CellReports* **15**, 2038-2049, doi:10.1016/j.celrep.2016.04.085 (2016).
- 5 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
- 6 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).
- 7 Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory domains. *Genome Research* **24**, 390-400, doi:10.1101/gr.163519.113 (2014).
- 8 Downen, J. M. *et al.* Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. *Cell* **159**, 374-387, doi:10.1016/j.cell.2014.09.030 (2014).
- 9 Gómez-Marín, C. *et al.* Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proceedings of the National Academy of Sciences* **112**, 7542, doi:10.1073/pnas.1505463112 (2015).
- 10 Hanssen, L. L. P. *et al.* Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nature Cell Biology* **19**, 952-961, doi:10.1038/ncb3573 (2017).
- 11 Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).
- 12 Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017-1021, doi:10.1126/science.1262088 (2015).
- 13 Symmons, O. *et al.* The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Developmental cell* **39**, 529-543, doi:10.1016/j.devcel.2016.10.015 (2016).
- 14 Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900-910, doi:10.1016/j.cell.2015.07.038 (2015).
- 15 de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Molecular Cell* **60**, 676-684, doi:10.1016/j.molcel.2015.09.023 (2015).
- 16 Paliou, C. *et al.* Preformed chromatin topology assists transcriptional robustness of Shh during limb development. *Proceedings of the National Academy of Sciences* **116**, 12390-12399, doi:10.1073/pnas.1900672116 (2019).
- 17 Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56, doi:10.1038/nature24281 (2017).
- 18 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e324, doi:10.1016/j.cell.2017.09.026 (2017).
- 19 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e922, doi:10.1016/j.cell.2017.05.004 (2017).
- 20 Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO Journal* **36**, 3573-3599, doi:10.15252/embj.201798004 (2017).
- 21 Rhodes, J. D. P. *et al.* Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell Reports* **30**, 820-835.e810, doi:<https://doi.org/10.1016/j.celrep.2019.12.057> (2020).

- 22 Kubo, N. *et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nature Structural & Molecular Biology* **28**, 152-161, doi:10.1038/s41594-020-00539-5 (2021).
- 23 Stik, G. *et al.* CTCF is dispensable for immune cell transdifferentiation but facilitates an acute inflammatory response. *Nature Genetics* **52**, 655-661, doi:10.1038/s41588-020-0643-0 (2020).
- 24 Thiecke, M. J. *et al.* Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Reports* **32**, 107929, doi:<https://doi.org/10.1016/j.celrep.2020.107929> (2020).
- 25 Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783, doi:10.1126/science.aau1783 (2018).
- 26 Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nature Genetics* **52**, 1151-1157, doi:10.1038/s41588-020-00716-8 (2020).
- 27 Luppino, J. M. *et al.* Cohesin promotes stochastic domain intermingling to ensure proper regulation of boundary-proximal genes. *Nature Genetics* **52**, 840-848, doi:10.1038/s41588-020-0647-9 (2020).
- 28 Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311, doi:10.1126/science.1067799 (2002).
- 29 Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).
- 30 Davies, J. O. J., Oudelaar, A. M., Higgs, D. R. & Hughes, J. R. How best to identify chromosomal interactions: a comparison of approaches. *Nature Methods* **14**, 125-134, doi:10.1038/nmeth.4146 (2017).
- 31 Hsieh, T.-H. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying chromosome conformation from the nucleosome to the entire genome. *Nature Methods* **13**, 1009-1011, doi:10.1038/nmeth.4025 (2016).
- 32 Hsieh, T.-Han S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108-119, doi:<https://doi.org/10.1016/j.cell.2015.05.048> (2015).
- 33 Hansen, A. S. *et al.* Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Molecular Cell* **76**, 395-411.e313, doi:<https://doi.org/10.1016/j.molcel.2019.07.039> (2019).
- 34 Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Molecular Cell*, doi:<https://doi.org/10.1016/j.molcel.2020.03.002> (2020).
- 35 Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Molecular Cell*, doi:<https://doi.org/10.1016/j.molcel.2020.03.003> (2020).
- 36 Hua, P. *et al.* Defining genome architecture at base-pair resolution. *Nature* **595**, 125-129, doi:10.1038/s41586-021-03639-4 (2021).
- 37 Li, Y. *et al.* CRISPR Reveals a Distal Super-Enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells. *PLOS ONE* **9**, e114485, doi:10.1371/journal.pone.0114485 (2014).
- 38 Vos, E. S. M. *et al.* Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression. *Molecular Cell*, doi:10.1016/j.molcel.2021.06.008.
- 39 Hsieh, T.-H. S. *et al.* Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1. *bioRxiv*, 2021.2007.2014.452365, doi:10.1101/2021.07.14.452365 (2021).
- 40 Liu, N. Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nature Genetics* **53**, 100-109, doi:10.1038/s41588-020-00744-4 (2021).
- 41 Valton, A.-L. *et al.* A cohesin traffic pattern genetically linked to gene regulation. *bioRxiv*, 2021.2007.2029.454218, doi:10.1101/2021.07.29.454218 (2021).
- 42 Kane, L. *et al.* Cohesin is required for long-range enhancer action. *bioRxiv*, 2021.2006.2024.449812, doi:10.1101/2021.06.24.449812 (2021).
- 43 Cuartero, S. *et al.* Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nature Immunology* **19**, 932-941, doi:10.1038/s41590-018-0184-1 (2018).



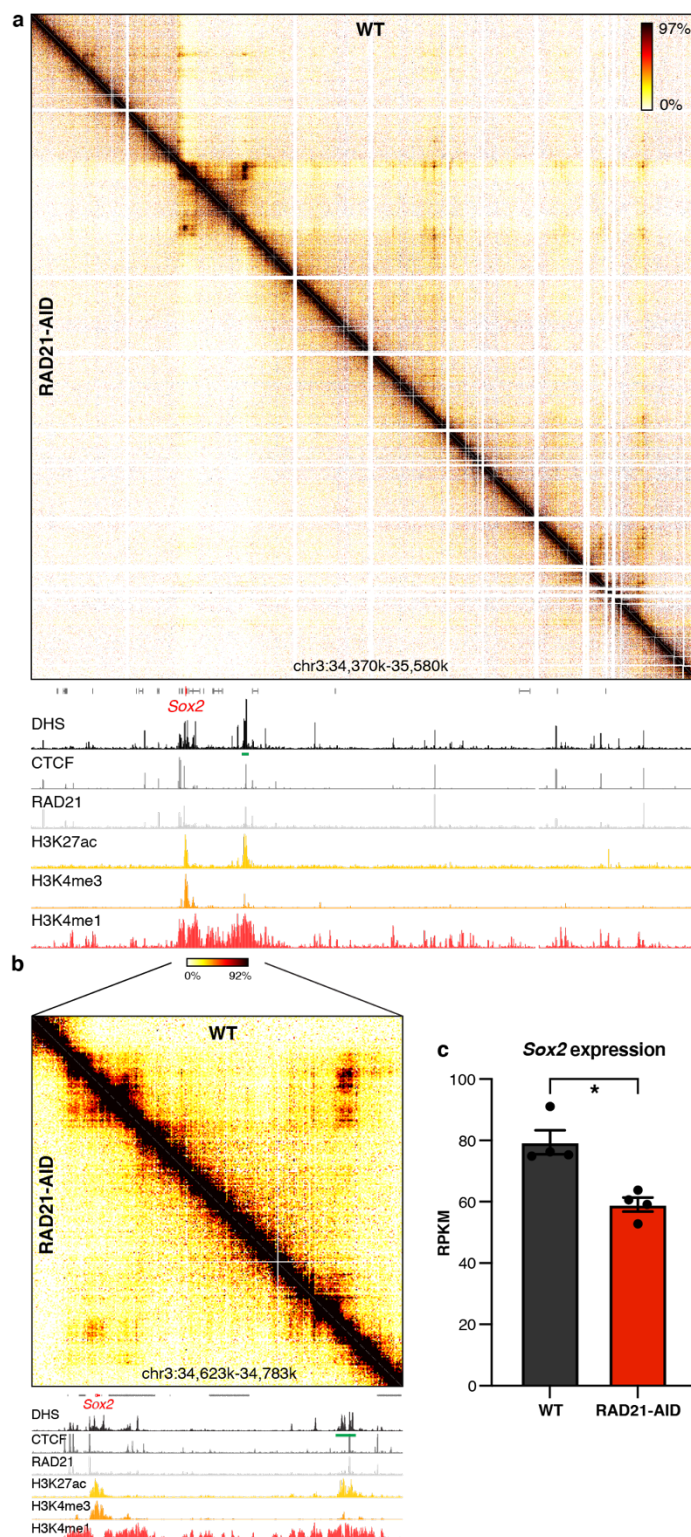
- 44 Calderon, L. *et al.* Activity-induced gene expression and long-range enhancer-promoter contacts in cohesin-deficient neurons. *bioRxiv*, 2021.2002.2024.432639, doi:10.1101/2021.02.24.432639 (2021).
- 45 Luan, J. *et al.* Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Reports* **34**, 108783, doi:<https://doi.org/10.1016/j.celrep.2021.108783> (2021).
- 46 Oudelaar, A. M. *et al.* Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nature Communications* **11**, 2722, doi:10.1038/s41467-020-16598-7 (2020).
- 47 Magoč, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)* **27**, 2957-2963, doi:10.1093/bioinformatics/btr507 (2011).
- 48 Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Research* **12**, 656-664 (2002).
- 49 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
- 50 Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* **9**, 999-1003, doi:10.1038/nmeth.2148 (2012).
- 51 Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology* **16**, 259, doi:10.1186/s13059-015-0831-x (2015).
- 52 Telenius, J., The, W. C. & Hughes, J. R. NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and CHIP-seq data mapping, high-resolution visualisation, and quality control. *bioRxiv*, 393413, doi:10.1101/393413 (2018).
- 53 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).
- 54 Whyte, W. A. *et al.* Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* **482**, 221-225, doi:10.1038/nature10805 (2012).
- 55 Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355-364, doi:10.1038/nature13992 (2014).

**Figure 1. Tiled-MCC generates local contact matrices with substantially improved sensitivity and resolution.**



**a.** Overview of the Tiled-MCC procedure. Cells are initially cross-linked with formaldehyde and permeabilized with digitonin. The chromatin is subsequently digested with MNase, which is followed by proximity ligation. After DNA extraction, the MCC libraries are sonicated and ligated with indexed sequencing adaptors. Multiplexed libraries are subsequently enriched for regions of interest using panels of biotinylated capture oligonucleotides, optimized for hybridization to MNase-digested 3C libraries. Since MNase does not cut at specific sites in the genome, the panels are designed to densely cover regions of interest with oligonucleotides of 70 nucleotides in length and an overlap of 35 nucleotides. This optimized capture oligonucleotide design results in ~80% of reads on target. **b-d.** Comparison of Tiled-MCC (top-right) and Micro-C (bottom-left) contact matrices at 500 bp resolution of the *Sox2* (**b**), *Prdm14* (**c**), and *Nanog* (**d**) loci in mES cells. Gene annotation (genes of interest in red, coding genes in black, non-coding genes in grey), DNase hypersensitive sites (DHS), and ChIP-seq data for CTCF are shown below the matrices. The axes of the DHS and CTCF ChIP-seq profiles are scaled to signal and have the following ranges; DHS: *Sox2* = 0-4.46, *Prdm14* = 0-6.45, *Nanog* = 0-10.25; CTCF: *Sox2* = 0-1833, *Prdm14* = 0-2168, *Nanog* = 0-3092. **e.** Density plots of ligation junctions identified by Tiled-MCC showing the micro-topology of *cis*-regulatory elements in the *Sox2* locus. Gene annotation, DNase hypersensitive sites (DHS), and ChIP-seq data for CTCF are shown below the plots. The axes of the DHS and CTCF ChIP-seq profiles are fixed and have the following ranges: DHS = 0-5; CTCF = 0-1500.

**Figure 2. Cohesin depletion results in reduced enhancer-promoter interactions in the *Sox2* locus.**

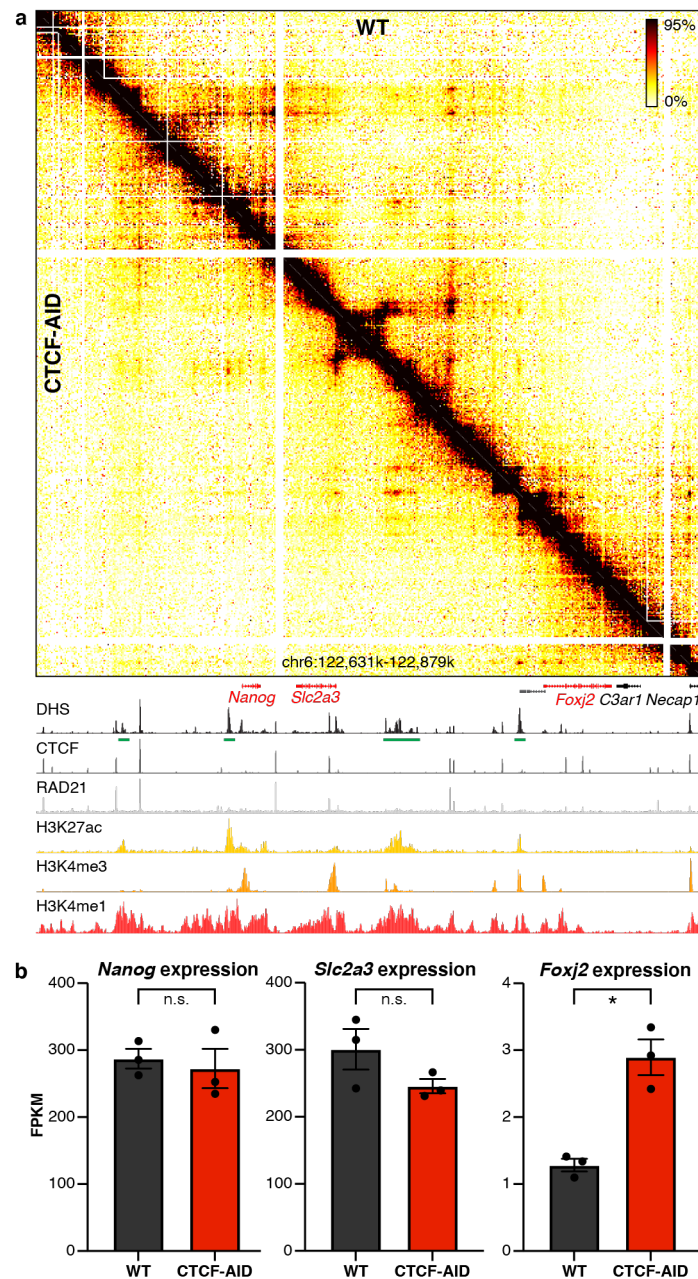


**a.** Tiled-MCC contact matrices of the *Sox2* locus in wild type (WT) mES cells (top-right) and auxin-treated RAD21-AID mES cells (bottom-left) at 500 bp resolution. Gene annotation (*Sox2* in red, coding genes in black, non-coding genes in grey), DNase hypersensitive sites (DHS) and ChIP-seq data for CTCF, Cohesin (RAD21), H3K27ac, H3K4me3 and H3K4me1 are shown below the matrices. The axes of the DHS and ChIP-seq profiles are scaled to signal and have the following ranges: DHS = 0-4.46;



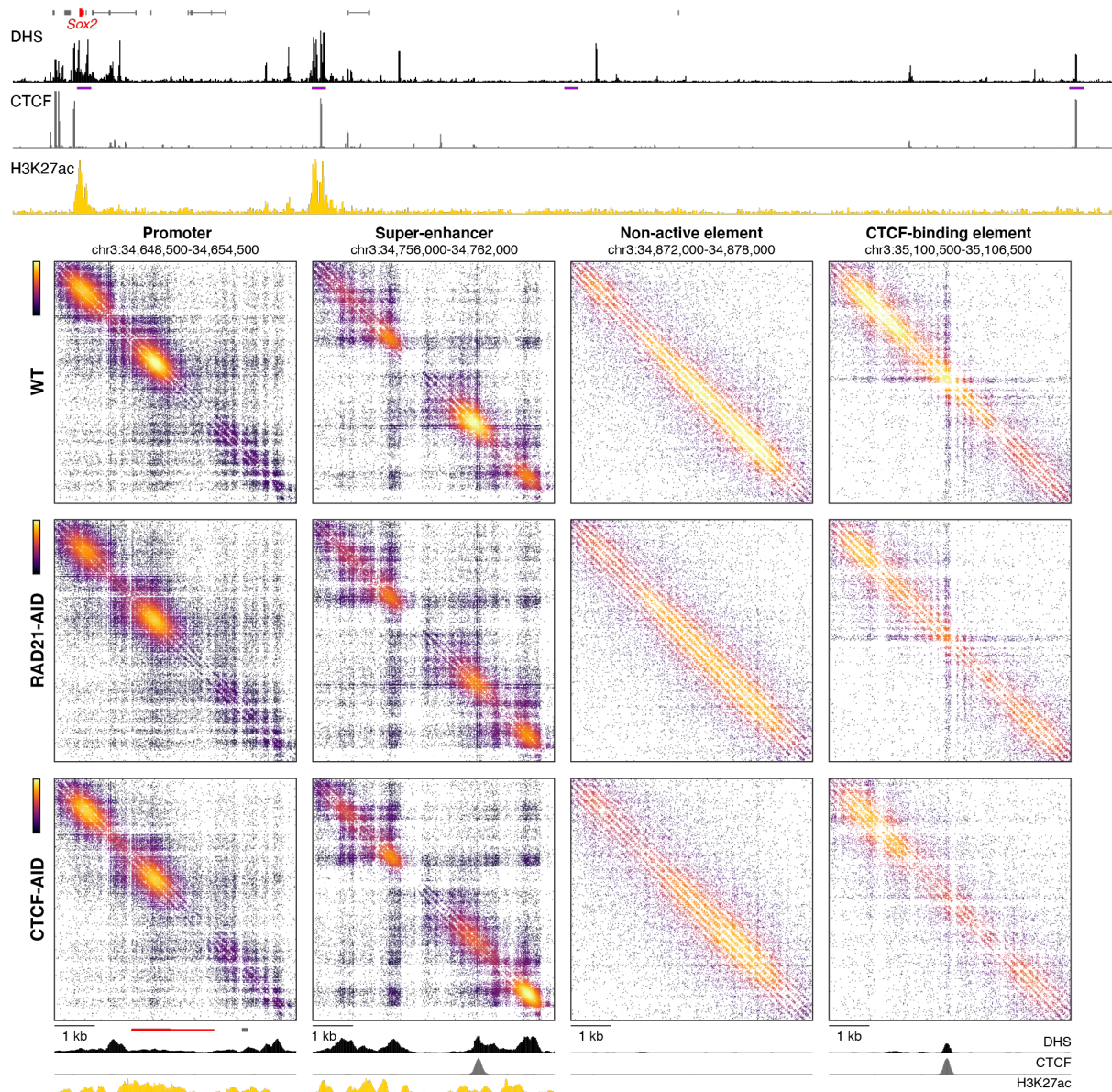
CTCF = 0-1833; RAD21 = 0-3318; H3K27ac = 0-48; H3K4me3 = 0-82; H3K4me1 = 0-1826. Enhancers of interest are indicated in green below the DHS profile. **b.** Zoomed view of the data presented in panel a to highlight the interactions between the *Sox2* promoter and its enhancers. The axes of the DHS and ChIP-seq profiles are scaled to signal and have the same ranges as in panel a, except for the CTCF ChIP-seq profile, which is scaled 0-300. **c.** Expression of *Sox2* in untreated (left) and auxin-treated (right) RAD21-AID mES cells, derived from RNA-seq data, normalized for reads per kilobase of transcript, per million mapped reads (RPKM). The bars represent the average of n=4 replicates and the error bars indicate the standard error of the mean. \* P = 4.46E-07.

**Figure 3. CTCF depletion results in ectopic enhancer-promoter interactions in the *Nanog* locus.**



**a.** Tiled-MCC contact matrices of the *Nanog* locus in wild type mES cells (top-right) and auxin-treated CTCF-AID mES cells (bottom-left) at 500 bp resolution. Gene annotation (genes of interest in red, coding genes in black, non-coding genes in grey), DNase hypersensitive sites (DHS) and ChIP-seq data for CTCF, Cohesin (RAD21), H3K27ac, H3K4me3 and H3K4me1 are shown below the matrices. The axes of the DHS and ChIP-seq profiles are scaled to signal and have the following ranges: DHS = 0-10.25; CTCF = 0-3092; RAD21 = 0-3414; H3K27ac = 0-58; H3K4me3 = 0-90; H3K4me1 = 0-2064. Enhancers of interest are indicated in green below the DHS profile. **b.** Expression of *Nanog*, *Slc2a3*, and *Foxj2* in untreated (left) and auxin-treated (right) CTCF-AID mES cells, derived from RNA-seq data, normalized for fragments per kilobase of transcript, per million mapped reads (FPKM). The bars represent the average of n=3 replicates and the error bars indicate the standard error of the mean. Significant (\*) and non-significant (n.s.) changes in expression are indicated. *Nanog*: P = 0.7030; *Slc2a3*: P = 0.1774; *Foxj2*: P = 0.0001.

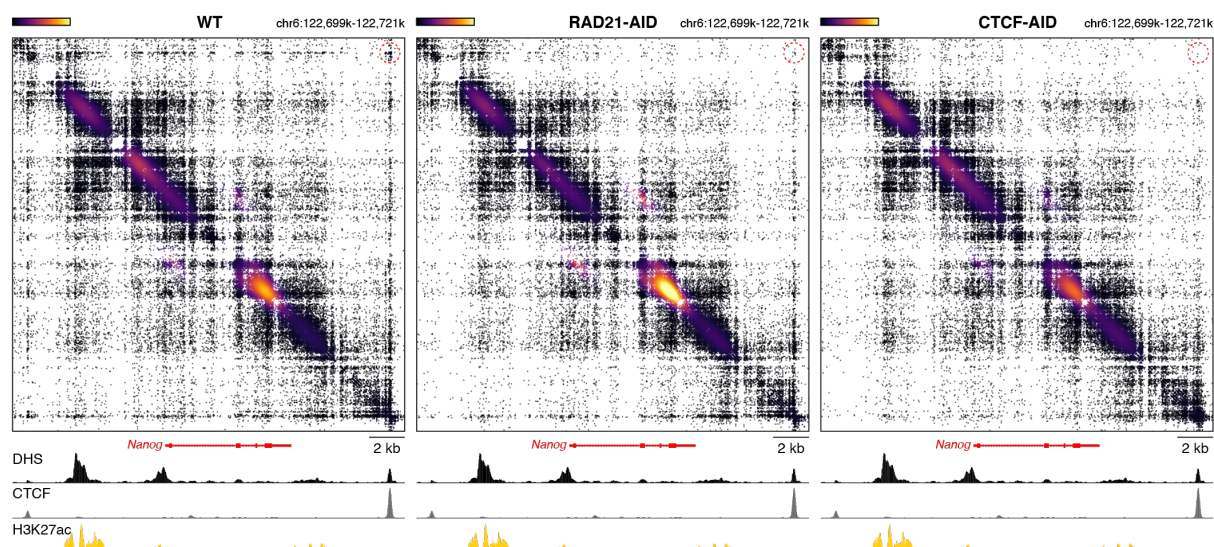
**Figure 4. Density plots of ligation junctions identified by Tiled-MCC reveal the micro-topology of *cis*-regulatory elements in the *Sox2* locus and their variable dependence on cohesin and CTCF.**



Density plots of ligation junctions identified by Tiled-MCC in the *Sox2* locus. Gene annotation (*Sox2* in red, coding genes in black, non-coding genes in grey), DNase hypersensitive sites (DHS), and ChIP-seq data for CTCF and H3K27ac for the extended *Sox2* locus are shown above the plots, with the regions covered in the density plots highlighted in purple. The density plots show a promoter, enhancer, non-active element and CTCF-binding element in order of the purple highlights and are annotated as described for the panel above. The axes of the DHS and ChIP-seq profiles for CTCF and H3K27ac are fixed and have the following ranges: DHS = 0-5; CTCF = 0-1500; H3K27ac = 0-50.



**Figure 5. The micro-topology of the *Nanog* locus upon cohesin and CTCF depletion.**



Density plots of ligation junctions identified by Tiled-MCC in the *Nanog* locus. Gene annotation, DNase hypersensitive sites (DHS), and ChIP-seq data for CTCF and H3K27ac are shown below the plots. The axes of the DHS and ChIP-seq profiles for CTCF and H3K27ac are fixed and have the following ranges: DHS = 0-8; CTCF = 0-2050; H3K27ac = 0-60. The interactions between the CTCF-binding elements are indicated with a red circle.