Region Capture Micro-C reveals coalescence of enhancers and promoters into nested microcompartments

Viraat Y. Goel[1,2,3], Miles K. Huseyin ⦿ [1,2,3] & Anders S. Hansen ⦿ [1,2,3] ✉

Although enhancers are central regulators of mammalian gene expression, the mechanisms underlying enhancer–promoter (E-P) interactions remain unclear. Chromosome conformation capture (3C) methods effectively capture large-scale three-dimensional (3D) genome structure but struggle to achieve the depth necessary to resolve fine-scale E-P interactions. Here, we develop Region Capture Micro-C (RCMC) by combining micrococcal nuclease (MNase)-based 3C with a tiling region-capture approach and generate the deepest 3D genome maps reported with only modest sequencing. By applying RCMC in mouse embryonic stem cells and reaching the genome-wide equivalent of ~317 billion unique contacts, RCMC reveals previously unresolvable patterns of highly nested and focal 3D interactions, which we term microcompartments. Microcompartments frequently connect enhancers and promoters, and although loss of loop extrusion and inhibition of transcription disrupts some microcompartments, most are largely unaffected. We therefore propose that many E-P interactions form through a compartmentalization mechanism, which may partially explain why acute cohesin depletion only modestly affects global gene expression.

3D genome structure regulates vital cellular processes including gene expression, DNA repair, genome integrity, DNA replication and somatic recombination[1,2]. Many insights into 3D genome structure have come from 3C assays, which have revealed structural hallmarks across at least three scales. First, active and inactive chromatin segregates into A- and B-compartments through a poorly understood compartmentalization mechanism[3,4]. Second, the genome is folded into loops and local domains called topologically associating domains (TADs) or loop domains[5–8] by loop-extruding cohesin complexes halted at CTCF boundaries[9,10]. Third, whereas A/B-compartments and TADs generally span hundreds to thousands of kilobases, recent work has hinted at finer-scale 3D chromatin interactions, including those linking enhancers and promoters[11–17]. Because enhancers are the primary units of gene expression control in mammals, there has been intense interest in resolving fine-scale E-P interactions; however, it has remained challenging to resolve fine-scale E-P interactions with current methods[8,18].

This challenge motivated us to develop a 3C method that effectively captures E-P interactions.

Advances in our understanding of 3D genome structure have been primarily driven by (1) deeper sequencing, (2) improved 3C protocols and (3) perturbation studies. First, A/B-compartments, TADs and loops were uncovered as deeper sequencing increased the number of captured unique contacts in 3C experiments from ~8 million[3] to ~450 million[5] to ~5 billion[7], respectively. Second, Micro-C overcomes the resolution limits imposed by Hi-C's dependence on restriction enzymes by digesting chromatin with MNase, which grants Micro-C nucleosome resolution and allows it to better resolve finer-scale regulatory interactions, including those between enhancers and promoters[8,11–13,15,19,20]. Third, perturbation studies have yielded profound mechanistic insights into 3D genome structure; for example, protein-depletion studies were pivotal in elucidating the roles of CTCF, cohesin and associated factors in the formation of TADs and loops[13,21–27].

[1]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Koch Institute for Integrative Cancer Research, Cambridge, MA, USA. ✉e-mail: ashansen@mit.edu

a

Overview of the RCMC workflow

b

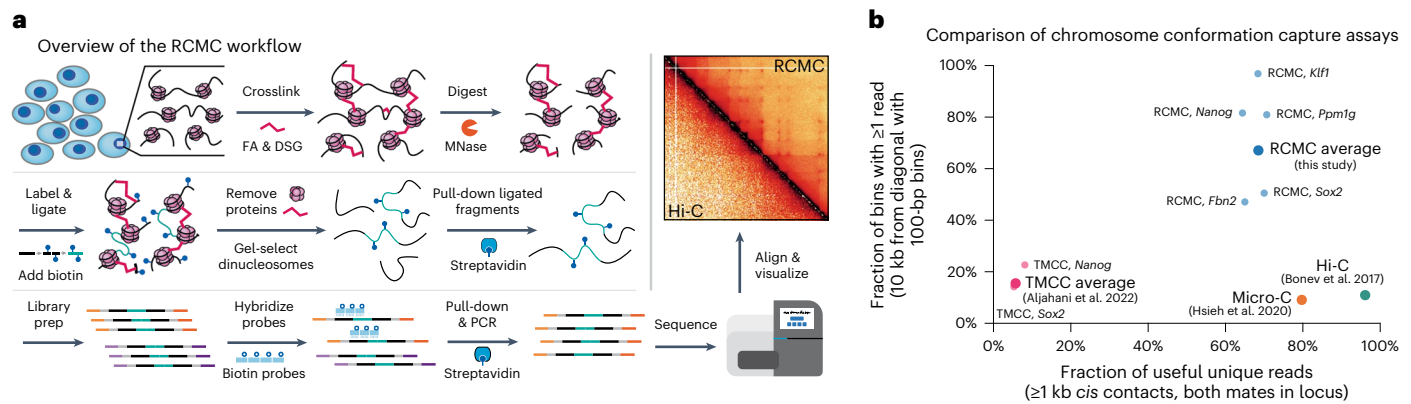Comparison of chromosome conformation capture assays

Fig. 1 | RCMC captures chromosome conformation at nucleosome resolution.
a, Overview of the RCMC protocol. Cells are chemically fixed, nuclei are digested with MNase and fragments are biotinylated, proximity-ligated, dinucleosomes gel-extracted and purified, library-prepped, PCR-amplified, and region-captured to create a sequencing library. After sequencing, mapping and normalization, the data are visualized as a contact matrix. **b**, Benchmarking comparison of RCMC against the highest-resolution Tiled-Micro-Capture-C (TMCC)[17], Micro-C[12]

and Hi-C[31] mESC datasets. Region-averaged calculations are shown for RCMC, TMCC, Micro-C and Hi-C, and calculations for individual captured regions are also shown for RCMC and TMCC. The *x* axis shows the fraction of all reads that uniquely map to the target region (both read mates fall within the captured region) that are structurally informative (*cis* contacts ≥1 kb). The *y* axis shows the fraction of all contact bins separated by 10 kb that contain at least one read at 100-bp resolution.

Nevertheless, despite decreasing sequencing costs, sequencing remains the key bottleneck for 3C assays. For a genome with *n* linear bins, sequencing costs to populate an n^2 pairwise contact matrix grow quadratically with *n*. For example, we estimate approximately $1.6 billion in sequencing costs alone to average one read per nucleosome-sized bin across the human genome (a total of $(3.3 \times 10^9$ bp/150 bp$)^2/2 = 2.4 \times 10^{14}$ reads). To overcome the prohibitive cost of sequencing inherent to current methods and facilitate the study of fine-scale 3D genome structure and EP interactions at loci of interest, we therefore sought to develop a 3C method that (1) strongly increases effective sequencing depth, (2) incorporates the latest advances in 3C-derived protocols and (3) is cost-effective for perturbation experiments.

Here, we address these three points by combining Micro-C with a tiling region capture approach[28,29] to enrich for entire regions of interest in a method we call Region Capture Micro-C (RCMC). We use RCMC to generate the deepest maps of 3D genome organization reported so far, achieving nucleosome resolution with a fraction of the sequencing of other methods. By reaching the local equivalent of ~317 billion unique contacts genome-wide, we discover patterns of previously unseen, fine-scale, focal and highly nested 3D interactions in gene-dense loci that we call microcompartments. Microcompartments frequently connect enhancers and promoters, and are largely robust to the loss of loop extrusion and inhibition of transcription, though some microcompartmental loops do change. Taken together, our results suggest that interactions between enhancers and promoters, now highly resolved by RCMC, may be driven by compartmentalization mechanisms rather than loop extrusion.

RCMC: development and benchmarking

To develop RCMC, we optimized the regular Micro-C protocol[11,12,15] to maximize library complexity and combined it with a tiling region capture approach[28,29] (Fig. 1a). Briefly, mouse embryonic stem cells (mESCs) were crosslinked with disuccinimidyl glutarate (DSG) and formaldehyde (FA) and digested to nucleosomes with MNase (Extended Data Fig. 1a,b), after which fragment ends were repaired with biotin-labeled nucleotides and then proximity ligated. After protein removal and reversal of crosslinks, we size-selected and pulled down ligated dinucleosomal fragments, and prepared a Micro-C sequencing library. Avoiding repetitive regions, we designed 80-mer biotinylated oligos tiling five regions of interest, each spanning between 425 kb and 1,900 kb (Extended Data Fig. 1c), and pulled down the tiled regions of interest

with 35–49% efficiency in a single step (Extended Data Fig. 1d). After paired-end sequencing and normalization[30] (Extended Data Fig. 1e), we obtained contact maps (Fig. 1a). To validate our RCMC contact maps, we compared them to high-resolution Hi-C[31] and Micro-C[12] for the same regions. Our RCMC data matched both Hi-C[31] and Micro-C[12] data at 2-kb resolution (Extended Data Fig. 1f), was reproducible (Extended Data Fig. 1g) and gave the expected contact frequency scaling (Extended Data Fig. 2a). Thus, RCMC captures all information in target regions obtained in prior multibillion contact studies[12,31].

Having validated RCMC, we next benchmarked it against other 3C datasets. Despite capturing ~2.6–3.3 billion unique contacts, the deepest Hi-C[31] and Micro-C[12] datasets in mESCs give sparse contact maps at fine (subkilobase) resolutions (Fig. 1b). In contrast, because RCMC focuses its sequencing reads in only regions of interest, almost all 100-bp-sized interaction bins showed at least one interaction for our most deeply sequenced region (*Klf1* Fig. 1b; Extended Data Fig. 2), and our RCMC maps matched genome-wide Micro-C[12] even after downsampling by ~100-fold (Extended Data Fig. 2d–f). Indeed, with relatively modest sequencing (Extended Data Fig. 2c) we captured the genome-wide equivalent of ~317 billion unique contacts at the *Klf1* region.

To visualize the improvements afforded by RCMC, we plotted contact maps comparing RCMC to Hi-C[31] and Micro-C[12] at our 5 captured regions (Extended Data Figs. 3–4). While A/B-compartments, TADs, and CTCF and cohesin-mediated structural loops are well-resolved in prior high-resolution Hi-C[31] and Micro-C[12] studies, resolving E-P interactions has proven more challenging[8,18]. To test the ability of RCMC to resolve E-P interactions, we captured a region around the *Sox2* gene and its regulatory elements (Fig. 2a). *Sox2* encodes a key pluripotency transcription factor, whose expression in mESCs is controlled by a well-characterized ~100-kb distal enhancer (*Sox2* control region (SCR))[32–34]. Although long-range *Sox2*-SCR interactions are visible in Hi-C and Micro-C, RCMC resolved the fine-scale substructure of the *Sox2*-SCR interactions; rather than one broad loop, *Sox2* forms multiple individual focal interactions with subelements of the SCR marked by Mediator binding and ATAC peaks (Fig. 2a). Furthermore, RCMC also revealed previously unobservable long-range interactions between a ~600–700 kb distal region near the *Fxr1* gene and *Sox2* and the SCR as well as strong compartmental exclusion of a ~550-kb intervening region (Extended Data Fig. 4a). Next, we focused on a ~300-kb segment of our most deeply sequenced region, the region around *Klf1* (Fig. 2b). Notably, RCMC revealed patterns of highly focal and nested interactions
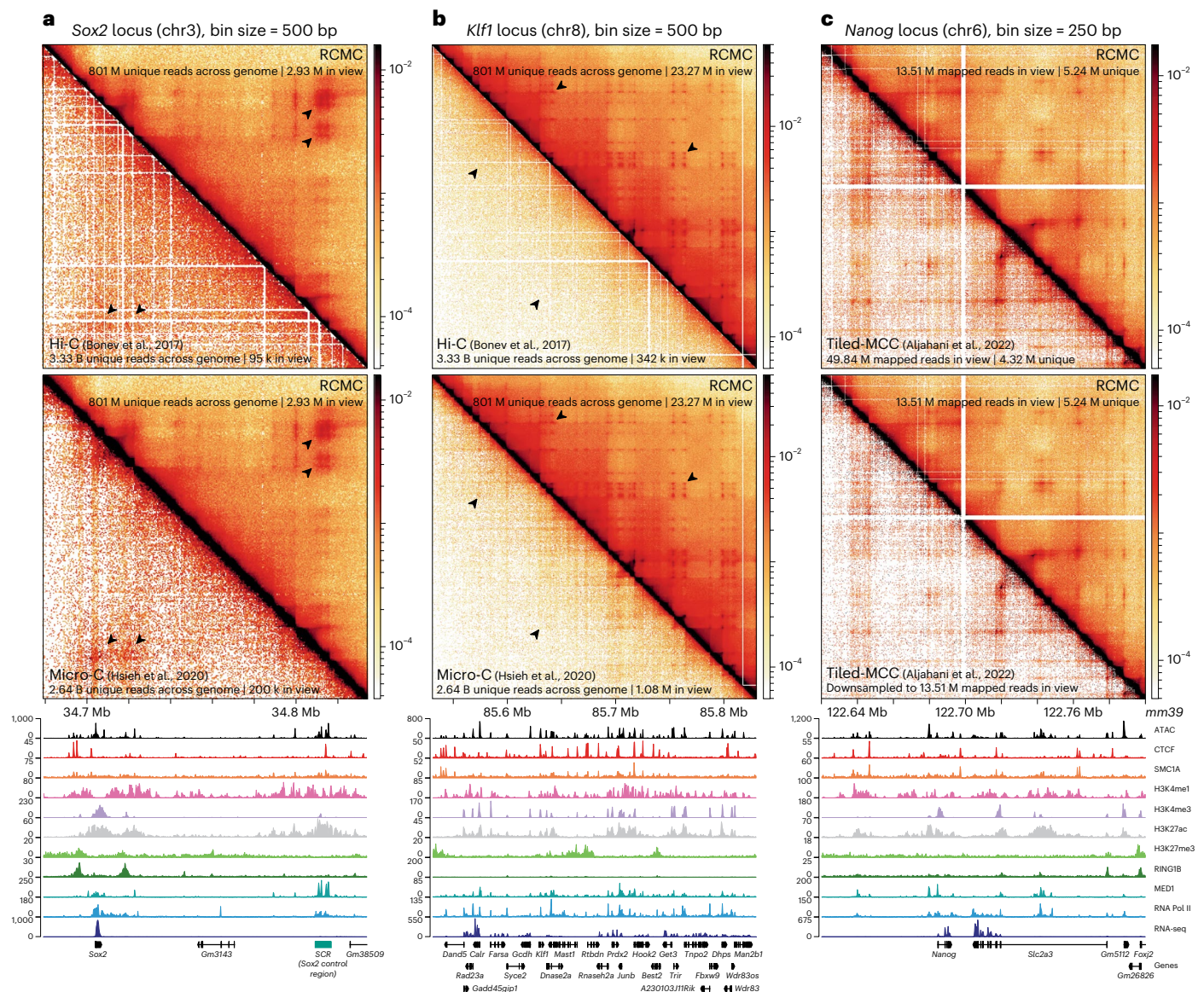
Fig. 2 | RCMC generates deep contact maps, reveals previously unresolved aspects of 3D genome structure, and outperforms other 3C methods.
a,b, Contact map comparison of RCMC against the deepest available mESC Hi-C (top; Bonev et al.[31]) and Micro-C (middle; Hsieh et al.[12]) datasets at the *Sox2* (a) and *Klf1* (b) regions at 500-bp resolution. Gene annotations and ATAC, chromatin immunoprecipitation with sequencing (ChIP-seq) and RNA sequencing (RNA-seq) (Supplementary Table 1) signal tracks are shown below the contact maps,

whereas the contact intensity scale is shown to the right. The RCMC data shown throughout this paper were pooled from two biological replicates in wild-type (WT) mESCs. **c**, Contact map comparison of RCMC against TMCC[17] at the *Nanog* locus at 250-bp resolution. Full datasets are visualized in the top contact map, and TMCC has been downsampled to match the total number of RCMC sequencing reads in view in the bottom contact map.

in the *Klf1* region that are not visible in genome-wide Hi-C or Micro-C data (Fig. 2b). We name these interactions microcompartments (see Discussion for rationale and definition). We conclude that for mapping genomic interactions within specific regions, RCMC outperforms genome-wide Hi-C and Micro-C at a fraction of the cost.

Finally, while our studies were ongoing, the related methods Micro-Capture-C (MCC)[16] and Tiled-Micro-Capture-C (TMCC)[17] were reported. Unlike RCMC, (T)MCC uses only formaldehyde for fixation[35], skips the pull-down of ligation products and the gel purification of dinucleosomes (Fig. 1a) and instead uses sonication to generate small fragments containing both ligated and unligated DNA. This allows (T)MCC to precisely sequence the ligation junction, which for RCMC requires longer-read sequencing. Thus, this affords (T)MCC base-pair resolution when capturing the interactions between regulatory elements[16,17]. However, by not enriching for the informative ligation

products, (T)MCC mainly captures unligated DNA fragments, resulting in most sequencing reads being uninformative (Fig. 1b). Indeed, with only slightly deeper sequencing, RCMC captured ~200 million unique >1-kb *cis* contacts in the target regions compared to just ~9–13 million for TMCC, underscoring the more than one order of magnitude higher efficiency of RCMC (Extended Data Fig. 2c). To directly compare RCMC to TMCC, we designed probes against the same *Nanog* region used in TMCC[17]. Due to the less efficient nature of TMCC, even with almost four-fold higher sequencing at the *Nanog* region, TMCC maps were noisier than RCMC, which became even more evident when we subsampled TMCC's sequencing depth to match RCMC (Fig. 2c and Extended Data Fig. 4b). In summary, we conclude that RCMC is more efficient for general 3D genome structure mapping of a region, whereas (T)MCC may be applied when it is necessary to resolve ligation junctions with base-pair resolution.
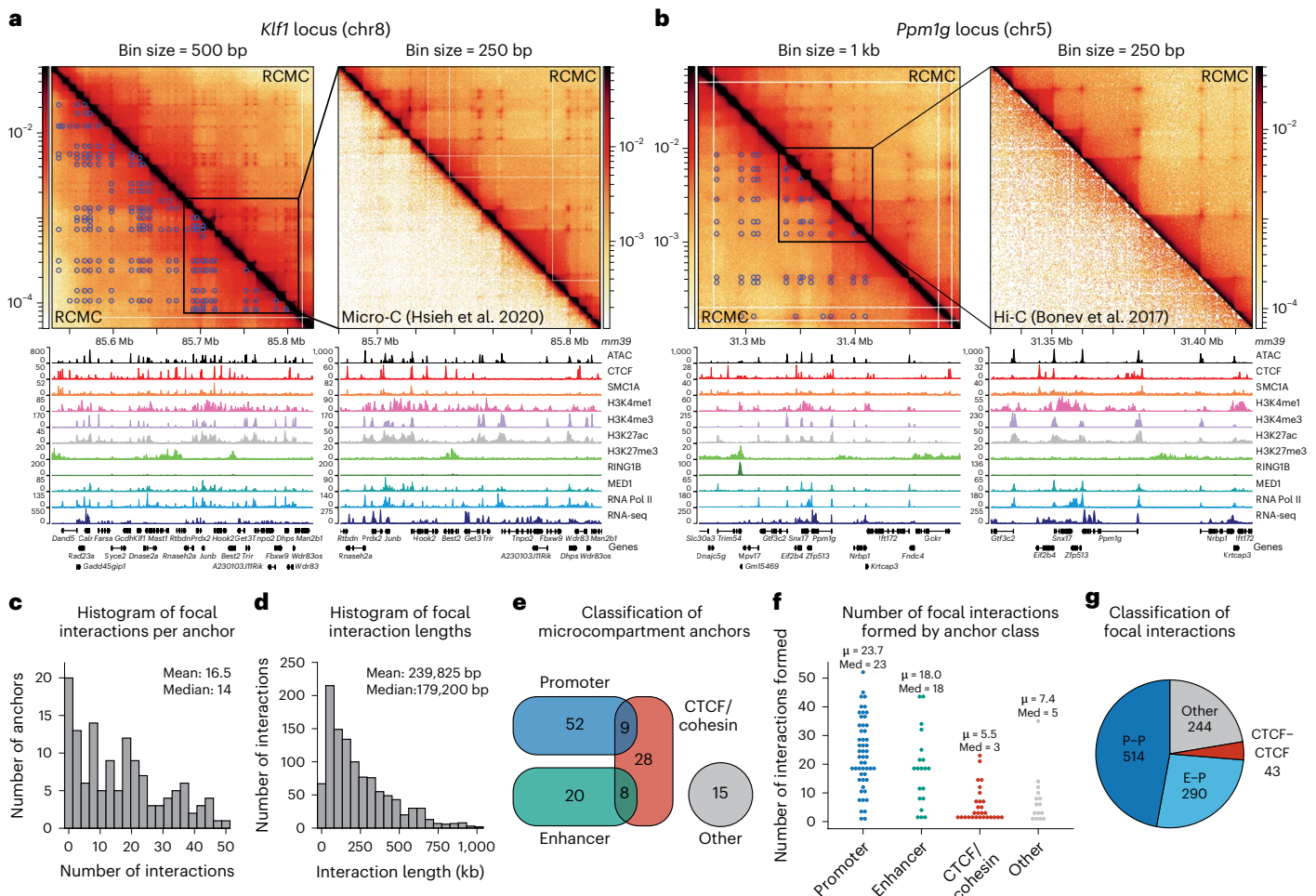
Fig. 3 | RCMC identifies highly nested focal interactions called microcompartments, which frequently connect enhancers and promoters. a,b, Contact map visualization of RCMC data and called microcompartments at the *Klf1* (a) and *Ppm1g* (b) locus at 500-bp (a) and 1-kb (b) resolution (left) and 250-bp resolution (zoom in, right). Manually annotated microcompartment contacts are shown below the contact map diagonal on the left, whereas comparisons against genome-wide Micro-C[12] (a) and Hi-C[31] (b) are shown on the right. **c,d,** Histograms showing distributions of the number of focal interactions formed by microcompartment anchors (c) and the lengths spanned by focal interactions in kilobases (d). **e,** Venn diagram of microcompartment anchor categories according to chromatin features overlapped by the anchor ±1 kb. Promoters were defined as regions around annotated transcription start

sites[51] ±2 kb, active enhancers as regions with overlapping peaks of H3K4me1 (ENCFF282RLA) and H3K27ac (GSE90893) in ChIP-seq data that did not overlap promoters, and CTCF/cohesin as regions with overlapping peaks of CTCF (GSE90994) and SMC1A (GSE123636) in ChIP-seq data. Other regions are those not overlapping any of these features. **f,** Swarm plot of the number of focal interactions formed by individual microcompartment anchors divided according to categories in panel e, including the mean (μ) and median (Med) for each distribution. Anchors fitting into more than one category were excluded. **g,** Fractions of loops classified into different categories: P–P (promoter-promoter), E–P, CTCF-CTCF (CTCF/cohesin-CTCF/cohesin) and other (other-other interactions, or any other combinations). CTCF-CTCF interactions do not include any anchors that overlap promoter or enhancer regions.

RCMC reveals nested focal interactions in gene-rich regions

RCMC data revealed highly nested and focal interactions in both the *Klf1* and *Ppm1g* regions, which were not visible in multibillion contact genome-wide Hi-C[31] and Micro-C[12] datasets (Figs. 2b and 3a,b and Extended Data Fig. 5a,b). We applied existing loop[36,37] and compartment calling algorithms[36] to identify these interactions, but they did not reliably detect them (Extended Data Fig. 5c). We therefore manually identified 132 anchors forming a total of 1,091 focal interactions in the gene-rich *Klf1* and *Ppm1g* regions (Fig. 3a,b and Extended Data Fig. 5d). Furthermore, we validated that these interactions were not due to incomplete contact map normalization[30] (Extended Data Figs. 1e and 6a) nor an artifact of increased accessibility at the anchors (only about half of all ATAC peaks result in 'dots,' and not all dots are anchored by ATAC peaks; Extended Data Fig. 6b–d).

Next, we observed that these interactions resemble both loops and compartments. Like loops, they give rise to focal enrichments

(dots in Fig. 3a,b) between two anchors and occasionally form contact domains as small as a few kilobases (squares in Fig. 3a,b). Like A/B-compartments, they result in nested, tessellated interactions in a checkerboard-like fashion, with a mean of ~17 interactions per anchor (mean interaction length: ~240 kb) and the most nested anchor forming 52 focal interactions (Fig. 3c,d). Because these highly nested and focal interactions (dots) resemble fine-scale compartmental interactions (Discussion), we refer to them as microcompartments.

To understand which genomic elements form microcompartments, we investigated the chromatin states of microcompartment anchors (Fig. 3c and Extended Data Fig. 7). About two-thirds of the identified microcompartment anchors overlapped either promoter (~46%) or enhancer (~21%) features (Fig. 3e and Extended Data Fig. 7), with the remaining anchors either corresponding to CTCF and cohesin-bound anchors or unknowns ('other'). Notably, however, promoters and enhancers formed many more focal interactions (Fig. 3f). Specifically, promoters and enhancers formed a mean of
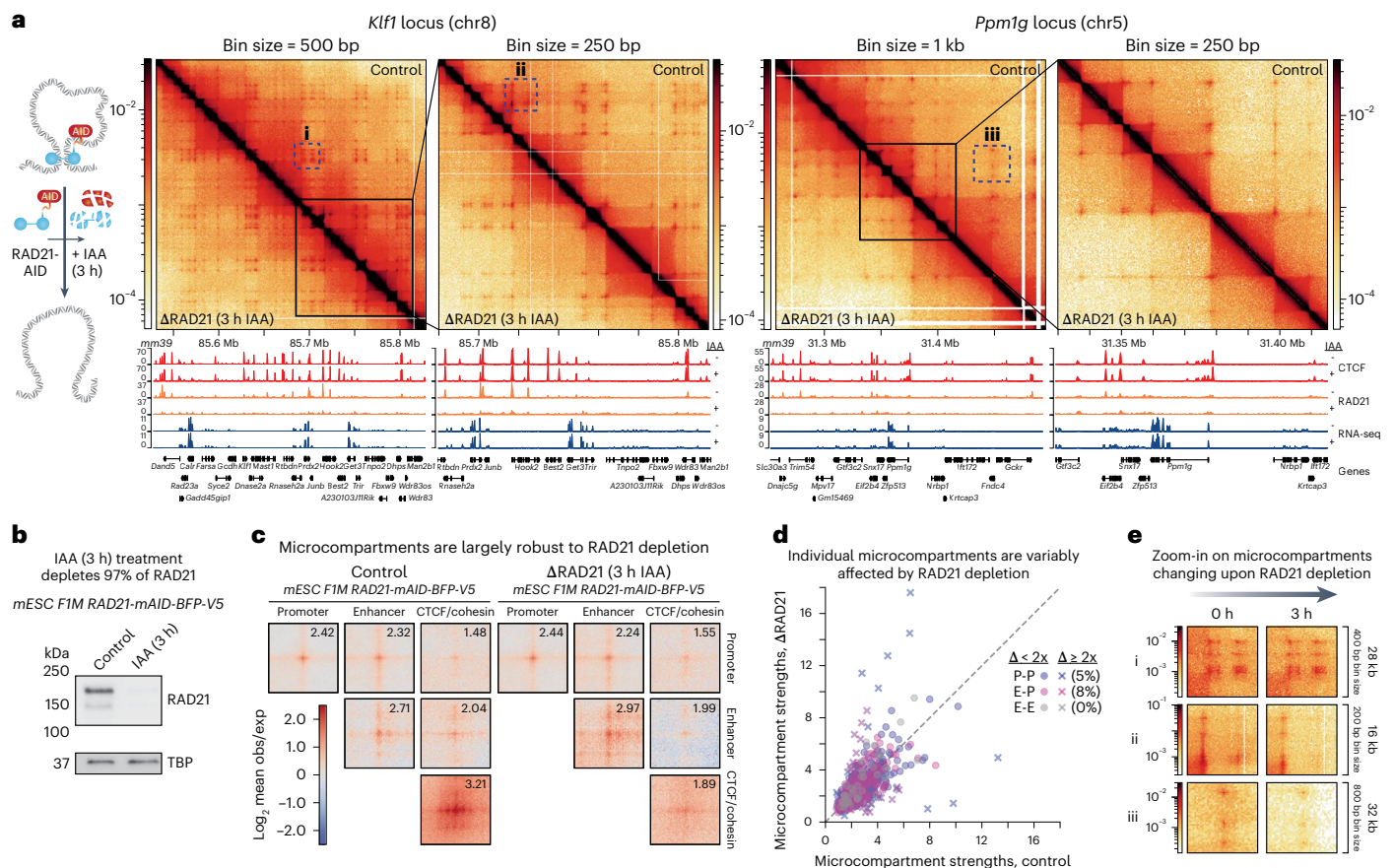
Fig. 4 | Most microcompartments are robust to the loss of loop extrusion.
a, Cohesin (RAD21) depletion does not strongly perturb most microcompartments. Left: Treatment paradigm for rapid depletion of RAD21 upon IAA treatment in clone F1M RAD21-mAID-BFP-V5 mESCs[13,38]. Right: Contact maps comparing DMSO-treated control (above) and RAD21-depleted (below) samples are shown for the *Klf1* and *Ppm1g* loci. **b**, Western blot showing near-complete (97%) depletion of RAD21 following 3 h IAA treatment. This western blot was performed once using cells collected simultaneously for RCMC. **c**, Aggregate peak analysis matrix of called microcompartmental contacts after RAD21 depletion compared to their respective controls, separated by the identity of each contact's constituent anchors. Plots show a 20-kb window centered

on the loop at 250-bp resolution. The background-normalized intensity for a $1,250 \times 1,250$ bp box around the central dot for each aggregate peak is shown in the upper right of each plot as a quantification of aggregate dot strength. **d**, Plot of individual microcompartment strengths (as quantified in panel c) in the RAD21-depleted (*y* axis) and control (*x* axis) conditions, shown for P-P (purple, $n = 418$), E-P (pink, $n = 238$) and E-E (gray, $n = 40$) loops. Interactions changing in strength by two-fold or more are visualized as x's, whereas interactions below the threshold are visualized as circles and percentages are noted. **e**, Zoomed-in contact maps of microcompartment examples in panel a that strengthen (i) or weaken (ii,iii) relative to the control treatment and the background in response to RAD21 depletion.

---

24 and 18 interactions, respectively, compared to just 5.5 and 7.4 for CTCF and cohesin, and 'other' anchors, respectively (Fig. 3f). Indeed, 74% of all annotated microcompartmental dots represented either P-P or E-P interactions, whereas only 4% of interactions were between anchors which exclusively overlapped CTCF and cohesin (Fig. 3g). Taken together, these observations suggest that microcompartments largely represent nested interactions between promoter and enhancer regions as well as some currently poorly understood 'other' regions.

Most microcompartments are robust to loss of loop extrusion

Having identified microcompartments as nested interactions frequently linking enhancers and promoters (Fig. 3a, b), we next took advantage of the cost-effective nature of RCMC to test the roles of loop extrusion and transcription (below) in forming these interactions.

First, we explored the role of cohesin and cohesin-mediated loop extrusion. Acute loss of cohesin strengthens large-scale A/B-compartments while simultaneously causing the global loss of TADs, loop domains and CTCF and cohesin-mediated structural loops[13,21,24,25,27,38]. Therefore, to understand whether cohesin regulates

microcompartments, we used our previously validated mESC line to acutely deplete the cohesin subunit RAD21 (mESC clone F1M RAD21-mAID-BFP-V5)[13,38] and performed RCMC across all five regions with and without 3 h of cohesin depletion (Fig. 4a and Extended Data Fig. 8a). The cohesin depletion was ~97% efficient (Fig. 4b), diminished the well-characterized CTCF and cohesin-mediated *Fbn2* loop[38] (Extended Data Fig. 8a), led to the expected change in contact frequency[21,23,24] (Extended Data Fig. 8b), and was reproducible between replicates (Supplementary Fig. 1), thus validating the cohesin depletion. As expected, the small fraction of interactions between CTCF and cohesin-bound sites showed large reductions in strength upon cohesin depletion (Fig. 4a,c and Extended Data Fig. 8a). However, the strengths of microcompartmental interactions, including E-P, E-E and P-P[13,17], were largely unaffected by cohesin depletion (Fig. 4c). Specifically, though we do see clear individual examples of especially P-P interactions that either slightly strengthen (Fig. 4e,i) or strongly weaken (Fig. 4e, ii-iii) after cohesin depletion (Fig. 4d,e), most microcompartmental interactions were largely unaffected (Fig. 4c). We therefore refine the microcompartment definition to interactions largely robust to cohesin depletion (see Discussion for full definition).

Most microcompartments are robust to loss of transcription

Second, we explored the role of transcription. We observed that microcompartments are largely formed between active promoter and enhancer regions (Fig. 3e,g and Extended Data Fig. 7), suggesting a relationship between active transcription and microcompartments. To understand if microcompartments are a downstream consequence of transcription, we abolished transcription by inhibiting transcription initiation by RNA polymerase II (Pol II) using triptolide. We chose two timepoints: 45 min, which was previously reported to modestly affect global E-P and P-P stripes[12], and 4 h, which was recently reported to greatly reduce punctate H3K4me3 (found at active promoters) and H3K27ac (found at active enhancers) marks in mESCs in addition to inhibiting transcription[39]. We performed RCMC across all five captured regions and ChIP-seq gave the expected reduction of RNA Pol II signal, with the 4-h triptolide treatment more thoroughly eliminating RNA Pol II at promoters and throughout gene bodies (Fig. 5a, b; Extended Data Fig. 9). We observed both weakened and strengthened E-P and P-P interactions (Fig. 5c–e), as well interesting dynamically changing interactions (for example, Fig. 5c,i increases in strength with 45 min of triptolide treatment but then weakens after 4 h). Nevertheless, the strong majority of microcompartmental interactions were largely unaffected by the inhibition of transcription (Fig. 5a,c–e). Our findings differ somewhat from recent studies reporting global weakening of E-P interactions after 14 h of ~80% depletion of RNA Pol II[20] or inhibition[19]. In addition to differences in cell type, treatment and treatment length, this difference may be due to the much lower depth (~1–1.7 billion Micro-C contacts)[19,20] used in these studies, which cannot resolve microcompartmental interactions and fine-scale E-P and P-P interactions (Fig. 3a and Extended Data Figs. 2e,f and 5b). Alternatively, because we only observe microcompartments in the very gene-dense *Klf1* and *Ppm1g* regions, prior findings[12,19,20] may apply more to individual, isolated E-P/P-P interactions instead of dense and nested microcompartments.

In summary, we conclude that microcompartments generally do not require transcription at short timescales and are more likely either independent from or formed upstream of transcription rather than forming as a downstream consequence of transcription.

Discussion

Here, we introduce RCMC as an accessible and affordable method for mapping 3D genome structure at unprecedented depth. Compared with Micro-Capture-C[16] methods such as TMCC[17], RCMC is much more efficient (Fig. 1b and Extended Data Fig. 2c), thus affording much higher depth with less sequencing. Another approach is to brute-force genome-wide Hi-C or Micro-C; by performing 150 separate Hi-C experiments and sequencing deeper than ever before, a recent study by Harris et al. reached 33 billion contacts[14]. However, such efforts[14] are expensive, not accessible to most labs, and poorly compatible with perturbation experiments vital to uncovering mechanisms of organization. Instead, with RCMC we reach the local equivalent of 317 billion contacts with relatively modest sequencing (Extended Data Fig. 2c). Thus, although genome-wide Micro-C may still be preferred for unbiased genome-wide 3D genome structure mapping, we propose RCMC as an ideal method for generating ultra-deep 3D contact maps and for perturbation experiments, albeit only for individual regions.

What molecular processes might drive microcompartment formation? Although cohesin-mediated loop extrusion is well established to generate focal interactions (loops)[9,10], microcompartmental loops are largely robust to acute cohesin removal and therefore likely not dependent on loop extrusion (Fig. 4a,c). Furthermore, although most microcompartmental loops connect enhancers and promoters, microcompartments are also generally robust to the acute loss of RNA Pol II transcription initiation (Fig. 5a,e). Instead, we propose that nested, multiway, and focal microcompartments correspond to small, punctate

A-compartments[14,40,41] that form through a compartmentalization mechanism, perhaps mediated by factors upstream of RNA Pol II initiation such as transcription factors and co-factors or active chromatin states[42]. Indeed, in the field of polymer physics, it is well established that block copolymers undergo microphase separation[4,43–45] when composed of distinct monomers that preferentially self-interact (Fig. 5f). Intuitively, if active chromatin regions at microcompartment anchors are selectively 'sticky' with each other, they will tend to co-segregate, resulting in the formation of nested, focal interactions (Fig. 5f). Microphase separation due to preferential interactions among active loci within a block copolymer might thus explain the formation of the striking pattern of interactions we observe (Figs. 3a,b and 5f). In summary, we tentatively define microcompartments as (1) highly nested, focal interactions that frequently connect promoters and enhancer regions often in gene-rich loci; (2) formed through a compartmentalization mechanism; and (3) for the most part independent of loop extrusion and transcription, at least on short timescales.

How do microcompartments compare to previously described 3D genome features? First, previous genome-wide Micro-C studies uncovered widespread short-range P-P and E-P links[12,13]. Similarly, many microcompartmental interactions connect promoters and enhancers. RCMC now better resolves these interactions, revealing them to be highly nested, frequently forming dozens of microcompartmental loops. Second, although differences in cell type preclude a direct comparison, the microcompartments described here also share features with the fine-scale A-compartment interactions recently described by Harris et al. that were proposed to segregate active enhancers and promoters into small A-compartments[14]. Indeed, examining the Hi-C data of Harris et al. at 1-kb resolution reveals structures with similarities to microcompartments, suggesting that microcompartments may be conserved to human cells (Extended Data Fig. 10). Further, along the lines of Harris et al., the microcompartments we observe form small contact domains, and their loops are more punctate as compared to CTCF and cohesin-mediated loops, which are more diffuse[14] (Figs. 4c and 5e).

Finally, our study provides insights into E-P interactions. Although some studies propose that cohesin is largely required for E-P interactions[27,46], others have suggested that cohesin is most important for very long-range[47–49] or inducible[48,50] E-P interactions or that cohesin is largely not required for the maintenance of E-P interactions[13,17]. Except for some CTCF and cohesin-bound enhancers and promoters, our data suggest that most P-P and E-P interactions are mediated by a compartmentalization mechanism distinct from loop extrusion. This may offer a mechanistic explanation for the observation that cohesin is not required for the short-term maintenance of most E-P interactions and that the effects of cohesin depletion on global gene expression are modest[13,17,25].

We end by noting some limitations and future directions. Although we can detect microcompartments with RCMC, further work will be necessary to fully understand their function and effect on gene expression. Furthermore, although we show that microcompartments are largely robust to loss of cohesin and transcription, further work is necessary to identify perturbation(s) that disrupt microcompartments. Subsequent studies will also be necessary to test our proposed compartmentalization mechanism of microcompartment formation, as well as to understand how microcompartments change during differentiation and across the cell cycle. Future imaging studies will also be required to understand the frequency and lifetime of microcompartmental interactions in live cells[38], as well as their multiway nature. Additionally, new computational tools will be required for automated microcompartment calling and analysis. Nevertheless, many of these questions can now be addressed with RCMC, and RCMC provides an accessible method to deeply resolve 3D genome structure in general and E-P interactions in particular across loci, cell types and disease states.
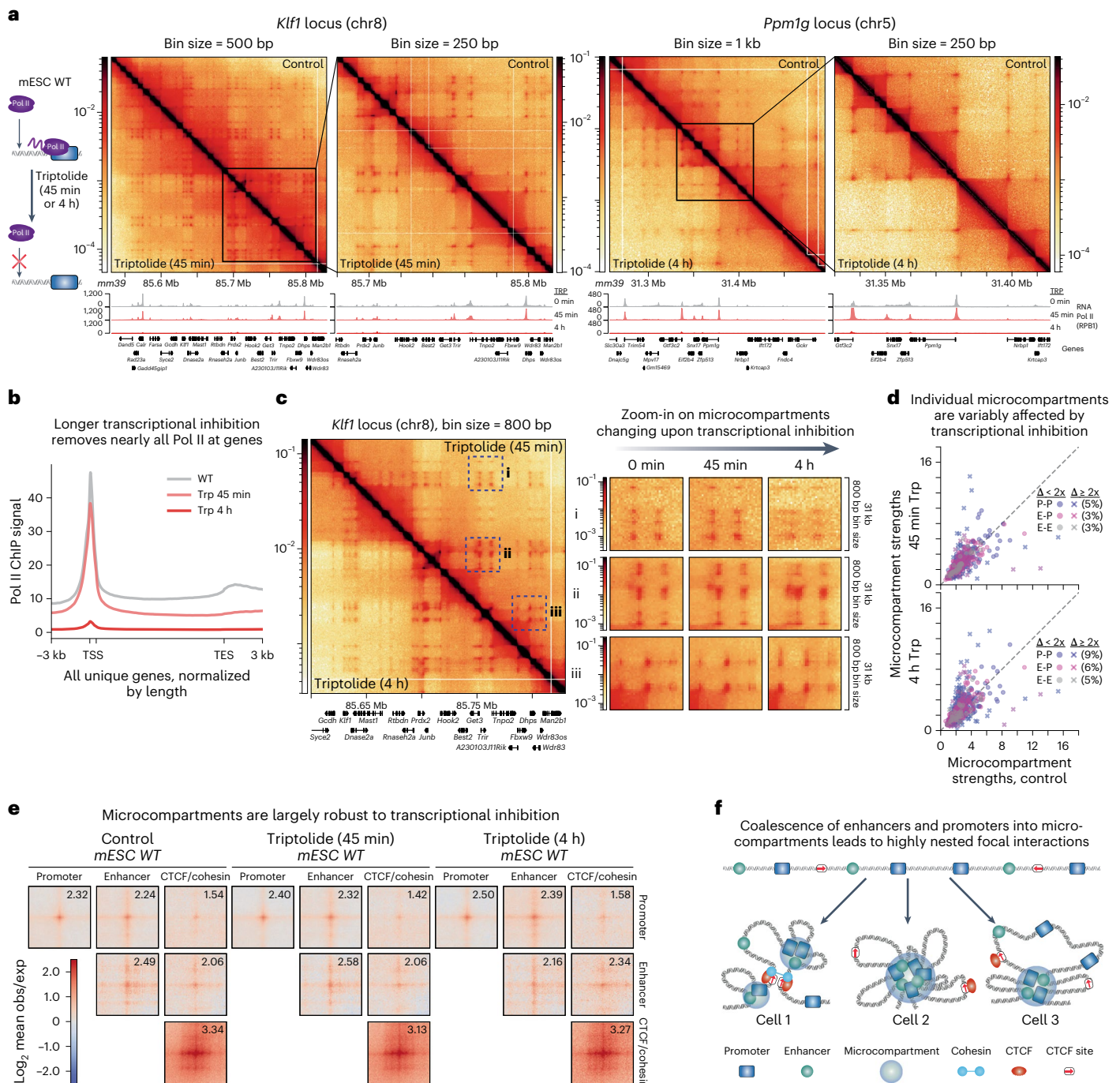
Fig. 5 | Most microcompartments are robust to the inhibition of transcription. a, Inhibition of transcription initiation with triptolide does not strongly affect most microcompartments. Left: Overview of triptolide treatment for WT mESCs (45 min or 4 h). Right: Contact maps comparing WT control (above) and transcriptionally inhibited (below) samples are shown for the *Klf1* locus (45-min timepoint shown vs. control) and the *Ppm1g* locus (4-hr timepoint shown vs. control). RNA Pol II ChIP-seq data (RPB1) are shown below. **b**, Aggregate RPB1 RNA Pol II ChIP-seq signal at genes after triptolide treatment (45 min and 4 h) and a control (WT). The *x* axis depicts all unique mouse genes normalized by length and flanked by 3 kb upstream and downstream of their transcription start site (TSS) and transcription end site (TES), respectively. The first 500 bp downstream of the TSS (marked by the second *x*-axis tick mark) are not normalized to avoid normalizing the core promoter against variable gene body lengths. **c**, Left: Contact maps comparing the transcriptional inhibition timepoints (45 min treatment above, 4 h treatment below) are shown for the *Klf1* locus. Right: Zoomed-in contact maps of microcompartments

across the control and triptolide treatment timepoints that weaken (i) or strengthen (ii,iii) in response to transcriptional inhibition. **d**, Plot of individual microcompartment strengths in the transcriptionally inhibited (*y* axis) and control (*x* axis) conditions, shown for P-P (purple, *n* = 418), E-P (pink, *n* = 238) and E-E (gray, *n* = 40) loops. Interactions changing in strength by two-fold or more are visualized as x's (percentages noted), and as circles otherwise. **e**, Aggregate peak analysis matrix of called microcompartmental contacts across the two transcriptional inhibition timepoints compared to the control, separated by the identity of each contact's constituent anchors. Plots show a 20-kb window centered on the loop at 250-bp resolution, with background-normalized dot intensities shown in the upper right of each plot. **f**, Proposed model for the formation of microcompartments. Coalescence of multiple promoters and enhancer elements in a gene-dense region may occur through A/B-block copolymer microphase separation, resulting in variable combinations of multiway interactions being present in different cells and giving rise to tessellated focal interactions in population-averaged RCMC data.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information, details of author contributions and competing interests and statements of data and code availability are available at https://doi.org/10.1038/s41588-023-01391-1.

References

1. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
2. Oudelaar, A. M. et al. The relationship between genome structure and function. *Nat. Rev. Genet.* **22**, 154–168 (2021).
3. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science.* **326**, 289–293 (2009).
4. Nuebler, J. et al. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl Acad. Sci. USA* **115**, E6697–E6706 (2018).
5. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
6. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
7. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
8. Goel, V. Y. et al. The macro and micro of chromosome conformation capture. *WIREs Dev. Biol.* **10**, e395 (2020).
9. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, 6456–6465 (2015).
10. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
11. Krietenstein, N. et al. Ultrastructural details of mammalian chromosome architecture. *Mol. Cell* **78**, 554–565 (2020).
12. Hsieh, T.-H. S. et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* **78**, 539–553 (2020).
13. Hsieh, T.-H. S. et al. Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, 1919–1932 (2022).
14. Harris, H. et al. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat. Commun.* (in the press).
15. Hansen, A. S. et al. Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. *Mol. Cell* **76**, 395–411 (2019).
16. Hua, P. et al. Defining genome architecture at base-pair resolution. *Nature* **595**, 125–129 (2021).
17. Aljahani, A. et al. Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat. Commun.* **13**, 2139 (2022).
18. Gasperini, M. et al. A Genome-wide framework for mapping gene regulation via cellular genetic screens. *Cell* **176**, 377–390 (2019).
19. Barshad, G. et al. RNA polymerase II and PARP1 shape enhancer-promoter contacts. Preprint at https://www.biorxiv.org/content/10.1101/2022.07.07.499190v1 (2022).
20. Zhang, S. et al. Enhancer–promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat. Genet.* https://doi.org/10.1038/s41588-023-01364-4 (2023).
21. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
22. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
23. Gassler, J. et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618 (2017).
24. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
25. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
26. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707 (2017).
27. El Khattabi, L. et al. A pliable mediator acts as a functional rather than an architectural bridge between promoters and enhancers. *Cell* **178**, 1145–1158 (2019).
28. Oudelaar, A. M. et al. Dynamics of the 4D genome during in vivo lineage specification and differentiation. *Nat. Commun.* **11**, 2722 (2020).
29. Jäger, R. et al. Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* **6**, 6178 (2015).
30. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
31. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
32. Zhou, H. Y. et al. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes Dev.* **28**, 2699–2711 (2014).
33. Li, Y. et al. CRISPR reveals a distal super-enhancer required for Sox2 expression in mouse embryonic stem cells. *PLoS One* **9**, e114485 (2014).
34. Chakraborty, S. et al. Enhancer–promoter interactions can bypass CTCF-mediated boundaries and contribute to phenotypic robustness. *Nat. Genet.* **55**, 280–290 (2023).
35. Akgol Oksuz, B. et al. Systematic evaluation of chromosome conformation capture assays. *Nat. Methods* **18**, 1046–1055 (2021).
36. Abdennur, N. et al. Cooltools: enabling high-resolution Hi-C analysis in Python. Preprint at https://www.biorxiv.org/content/10.1101/2022.10.31.514564v1 (2022).
37. Roayaei Ardakany, A. et al. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol.* **21**, 256 (2020).
38. Gabriele, M. et al. Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science.* **376**, 496–501 (2022).
39. Wang, Z. et al. Prediction of histone post-translational modification patterns based on nascent transcription data. *Nat. Genet.* **54**, 295–305 (2022).
40. Rosencrance, C. D. et al. Chromatin hyperacetylation impacts chromosome folding by forming a nuclear subcompartment. *Mol. Cell* **78**, 112–126 (2020).
41. You, Q. et al. Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution. *Nat. Biotechnol.* **39**, 225–235 (2021).
42. Rippe, K. et al. Functional organization of RNA polymerase II in nuclear subcompartments. *Curr. Opin. Cell Biol.* **74**, 88–96 (2022).
43. Leibler, L. Theory of microphase separation in block copolymers. *Macromolecules* **13**, 1602–1617 (1980).
44. Meier, D. J. Theory of block copolymers. I. Domain formation in A-B block copolymers. *J. Polym. Sci. Part C. Polym. Symp.* **26**, 81–98 (1969).
45. Fujishiro, S. et al. Generation of dynamic three-dimensional genome structure through phase separation of chromatin. *Proc. Natl Acad. Sci. USA* **119**, e2109838119 (2022).

46. Thiecke, M. J. et al. Cohesin-dependent and -independent mechanisms mediate chromosomal contacts between promoters and enhancers. *Cell Rep.* **32**, 107929 (2020).

47. Kane, L. et al. Cohesin is required for long-range enhancer action at the Shh locus. *Nat. Struct. Mol. Biol.* **29**, 891–897 (2022).

48. Calderon, L. et al. Cohesin-dependence of neuronal gene expression relates to chromatin loop length. *Elife* **11**, e76539 (2022).

49. Rinzema, N. J. et al. Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat. Struct. Mol. Biol.* **29**, 563–574 (2022).

50. Cuartero, S. et al. Control of inducible gene expression links cohesin to hematopoietic progenitor self-renewal and differentiation. *Nat. Immunol.* **19**, 932–941 (2018).

51. Navarro Gonzalez, J. et al. The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res.* **49**, D1046–D1057 (2021).

Methods

Experimental procedure

Overview of the RCMC experiment. RCMC was developed by merging Micro-C[12] with tiling region capture of a locus[28,29]. An overview of the RCMC protocol is provided below, and a detailed protocol is provided as Supplementary Information. The data generated in this paper come from merging of two RCMC biological replicates for each of the five tested conditions (WT, transcriptional inhibition for 45 min or for 4 h, cohesin depletion and a cohesin depletion control). For four of the tested conditions (all except transcriptional inhibition for 4 h), the first biological replicate is a compilation of three technical replicates generated from the same batch of harvested cells. Biological replicates were generated by harvesting (culturing, crosslinking, aliquoting and snap-freezing) 125–200 M cells for each tested condition, after which downstream RCMC steps (Micro-C and Capture) were applied to five snap-frozen 5 M cell aliquots to generate one to three technical replicates for each biological replicate. Reaction volumes for performing RCMC on both 1 M and 5 M cell samples are provided in the Supplementary Protocol.

Cell culture. mESCs (JM8.N4 mESCs[52]; Research Resource Identifier: RRID:CVCL_J962; obtained from the KOMP Repository at UC Davis) were cultured at 37 °C with 5% CO_2 on plates coated with 0.1% gelatin solution (Sigma-Aldrich #G1890) under feeder-free conditions in medium consisting of KnockOut DMEM (ThermoFisher #10829-018) with 15% FBS (HyClone, SH30396.03, lot no. AE28209315) and 1,000 U ml^{-1} LIF (homemade[53]), 1 mM MEM Non-Essential Amino Acid Solution (ThermoFisher #11140-050), 2 mM GlutaMAX (ThermoFisher #35050061) 100 µg ml^{-1} penicillin-streptomycin (ThermoFisher #15140-122) and 0.1 mM 2-mercapoethanol (ThermoFisher #31350010) supplemented with 2i, 10 µM MEK inhibitor (Tocris #PD0325901) and 3 µM GSK inhibitor (Sigma-Aldrich #SML1046). F1M RAD21-mAID-BFP-V5 JM8.N4 mESCs were previously generated and validated in the laboratory[38]. mESCs were fed daily by replacing half of the medium and passaged every 2 days with TrypLE Express Enzyme (ThermoFisher #12605036). One day before treatment and harvesting, cells were swapped to medium as described above without 2i.

HEK293T cells were obtained from ATCC (CRL-3216) and were cultured at 37 °C with 5% CO_2 in DMEM supplemented with 10% FBS, 2 mM L-glutamine, 1× penicillin-streptomycin and 0.5 mM β-mercaptoethanol.

Depletion of cohesin. Depletion of cohesin was achieved using indole-3-acetic acid (IAA) treatment of the cell line clone F1M RAD21-mAID-BFP-V5 as previously described[13,38]. A 250 mM IAA (BioAcademia #30-003-10) stock was prepared by dissolving the drug in DMSO. F1M RAD21-mAID-BFP-V5 JM8.N4 mESCs[38] were grown to ~80% confluency in medium as described above, with a swap to 2i-free medium 24 h before treatment. Cells were washed once with PBS and fed fresh 2i-free medium containing either only DMSO (untreated control) or 500 µM IAA (cohesin depleted), incubated for 3 h and then harvested.

Inhibition of transcription. Inhibition of RNA Pol II activity was achieved using triptolide treatment as previously described[12]. A 1 mM triptolide (Sigma-Aldrich #T3652) stock was prepared by dissolving the drug in DMSO. WT JM8.N4 mESCs[52] were grown to ~80% confluency in medium as described above, with a swap to 2i-free medium 24 h before treatment. Cells were washed once with PBS and fed fresh 2i-free medium containing 1 µM triptolide, incubated for 45 min or for 4 h and then harvested.

Crosslinking. Cells were doubly crosslinked to fix protein-protein and protein-DNA interactions using DSG (disuccinimidyl glutarate, 7.7 Å) (ThermoFisher #20593) and formaldehyde (ThermoFisher #28906), respectively. Crosslinking medium was prepared by diluting freshly made DSG stock solution (300 mM DSG in DMSO) to 3 mM in 1× PBS (ThermoFisher #10010031). Trypsinized cells were resuspended to single cells, counted, washed in PBS and then resuspended in crosslinking medium at a concentration of 1 M cells ml^{-1}. The crosslinking reaction was gently mixed at room temperature for 35 min, after which formaldehyde was added to a final concentration of 1%. The double crosslinking reaction was mixed at room temperature for an additional 10 min before quenching with Tris buffer pH 7.5 (K-D Medical #RGE-3370) at a final concentration of 0.375 M. Treatments for non-WT samples (1 µM triptolide, 500 µM IAA or DMSO) were added to all harvesting reagents used before Tris quenching (PBS, trypsin, trypsin-quenching media and crosslinking medium) to avoid post-treatment rescue during the crosslinking process. Crosslinked cells were washed twice with 1× PBS, recounted to quantify any sample loss during fixation and then partitioned into 5 M cell aliquots that were pelleted and snap-frozen in liquid nitrogen for storage at −80 °C.

MNase titration. Digesting the crosslinked genome to the nucleosome-sized fragments (150–200 bp) necessary to capture nucleosome-resolution DNA contacts requires a titration to identify the ideal MNase digestion concentration and reaction conditions. Accordingly, MNase titrations were performed for each batch of crosslinked cells before performing the RCMC protocol. The titration involved MNase digestion of 1 M or 5 M cell samples varying MNase concentrations, reversal of crosslinks, DNA purification and gel-based separation to visualize the distribution of fragment sizes (see corresponding sections below). Ideal digestion concentrations were identified by samples digested to primarily (~80%) mononucleosomal fragments (150–200 bp), few (~15–20%) dinucleosomal fragments (250–350 bp) and a faint but visible band (<5%) of trinucleosomal fragments (400–500 bp) (Extended Data Fig. 1a).

MNase digestion. Cell membranes were solubilized to extract intact nuclei by resuspending crosslinked 5 M cell pellets in Micro-C Buffer #1 (MB#1; 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 5 mM $MgCl_2$, 1 M $CaCl_2$, 0.2% NP-40 Alternative (Millipore Sigma-Aldrich #492018), 1× Protease Inhibitor Cocktail (Sigma-Aldrich #5056489001)) at 1 M cells per 100 µl for 20 min on ice. Following an MB#1 wash, samples were resuspended in 100 µl MB#1 and the ideal amount of 20 U µl^{-1} MNase (Worthington Biochem #LS004798) determined by the MNase titration was added. This digestion reaction was mixed at 37 °C for 20 min on a thermomixer before being quenched with 4 mM EGTA (bioWORLD #40520008) and heat inactivated at 65 °C for 10 min. Digested nuclei were washed twice with ice-cold Micro-C Buffer #2 (50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 100 µg ml^{-1} BSA (Sigma-Aldrich #B8667)).

End repair and labeling. To generate blunt ends on digested DNA fragments before proximity ligation and add biotinylated nucleotides, a series of enzymatic processing steps were performed. First, to catalyze the addition of 5′-phosphate groups and the removal of 3′-phosphate groups, digested samples generated from 5 M cell inputs were incubated in end-repair reactions (50 U T4 Polynucleotide Kinase (New England BioLabs #M0201), 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 100 µg ml^{-1} BSA, 2 mM ATP (ThermoFisher #R1441) and 5 mM DTT (Sigma-Aldrich #10197777001), in water) at 37 °C for 15 min while mixing. To create 5′ fragment overhangs for end-blunting and labeling, 50 U DNA Polymerase I Klenow Fragment (New England BioLabs #M0210) was added to the reaction and incubated at 37 °C for 15 min while mixing. Next, a mixture of dNTPs in end-labeling buffer (66 µM each of dTTP (Jena Bioscience #NU-1004), dGTP (Jena Bioscience #NU-1003), biotin-dATP (Jena Bioscience #NU-835-BIO14), and biotin-dCTP (Jena Bioscience #NU-809-BIOX), 1× T4 DNA Ligase Buffer, 100 µg/ml BSA, in water) was added to the reaction. This reaction was incubated at room temperature for 45 min with interval mixing before being quenched by 30 mM EDTA (Invitrogen #15575020)

and heat inactivated at 65 °C for 20 min. Finally, end-blunted and biotin-labeled nuclei were washed once with Micro-C Buffer #3 (50 mM Tris-HCl pH 7.5, 10 mM MgCl$_2$ and 100 µg ml^{-1} BSA).

Proximity ligation and removal of unligated biotin. Proximity ligation was performed by incubating labeled chromatin in a ligation reaction (10,000 U T4 DNA Ligase (New England BioLabs #M0202), 1× T4 DNA Ligase Buffer, 100 µg ml^{-1} BSA, in 500 µl water) at room temperature for at least 2.5 h with gentle mixing. To remove biotinylated dNTPs from all unligated fragment ends, samples were digested by 1,000 U Exonuclease III (New England BioLabs #M0206) in reaction buffer (1× NEBuffer #1 in water) at 37 °C for 15 min with interval mixing.

DNA purification and size-selection. To prepare ligated DNA for library generation, DNA was reverse crosslinked and proteins and RNA were digested by adding 1% SDS (Sigma-Aldrich #L3771), 2 mg ml^{-1} Proteinase K (Viagen Biotech #501-PK), 250 mM NaCl and 100 µg ml^{-1} RNaseA (ThermoFisher #EN0531) to the samples and incubating at 65 °C overnight. DNA was extracted using phenol:chloroform:isoamylic alcohol (Sigma-Aldrich #P2069) in a 1:1 volumetric ratio using 5PRIME Phase Lock Gel Light tubes (Quantabio #2302820). The aqueous phase was further purified using the Zymo DNA Clean & Concentrator kit (Zymo Research #D4034) according to the kit manual.

Dinucleosome-sized DNA fragments (250–350 bp) were isolated by extraction from a 1% agarose gel (VWR #97062) (Extended Data Fig. 1b). Gel extracts were purified using the Zymo Gel Purification kit (Zymo Research #D4008), and samples were quantified by Qubit 1× dsDNA High Sensitivity Assay (Invitrogen #Q33231). Sample ends were polished and blunted again using the End-It enzyme reaction (Lucigen #ER81050) at 25 °C for 45 min, followed by reaction inactivation at 65 °C for 10 min.

Ligated DNA contact fragments were isolated by pulling down biotin-bound fragments using Dynabeads MyOne Streptavidin T1 (Invitrogen #65601). DNA samples were bound to beads in a Binding and Wash Buffer (1 M NaCl, 5 mM Tris-HCl pH 7.5, 500 µM EDTA, 0.1% Tween-20 (Sigma-Aldrich #P8074)) at room temperature for at least 30 min with mixing. After two washes with the Binding and Wash Buffer, the bead-bound samples were washed once with 10 mM Tris-HCl pH 7.5 before library preparation.

Library preparation. Illumina library preparation was performed using the NEBNext Ultra II kit (New England BioLabs #E7645) to end-repair, A-tail, and adaptor ligate the bead-bound samples. All steps were performed as directed by the manual, except that incubations included interval shaking (1 min on, 3 min off) at 1,000 rpm. Sample washes were performed using Binding and Wash Buffer and 10 mM Tris-HCl pH 7.5 washes. To determine the minimum number of PCR cycles to meet input material guidelines for capture or sequencing, a test library amplification was performed with 5% or less of the prepped library to quantify the yield. The test PCR reaction mixture was run on an agarose gel and yield was quantified using image quantification software Image Studio Lite (LI-COR Biosciences). Ten or fewer PCR cycles to meet capture input requirements is optimal to reduce PCR duplicates, and the RCMC replicates in this paper used seven to eight PCR cycles for final library amplification. All library amplifications were done using sequencing indices from the NEB Multiplex Oligos for Illumina Primer Set 1 (New England BioLabs #E7335) and the KAPA HiFi HotStart ReadyMix enzyme (Roche #07958927001). Following library amplification, the T1 Dynabeads containing the original bead-bound samples were removed and the amplified libraries were purified to remove adaptor dimers, primers and contaminants using AmPure XP beads (Beckman Coulter #A63880). Purified libraries were quantified via Fragment Analyzer and qPCR at the MIT BioMicro Center to determine library concentrations for pooling before capture.

Capture probe design. Target loci of interest were identified based on genomic features or E-P relationships of interest. *Klf1* and *Ppm1g* were selected as gene-rich loci, *Fbn2* was selected as a gene-poor control with a well-established CTCF- and cohesin-mediated loop[38], and *Sox2* and *Nanog* were later selected as loci for comparing RCMC against TMCC (Extended Data Fig. 3). Using the UCSC Genome Browser and HiGlass visualization of existing mESC 3C datasets, locus bounds were selected to include visible local structures and genomic features in roughly 1-Mb-sized regions. Once loci had been selected, 80-mer probes were designed to tile end-to-end without overlap across the capture loci through Twist Bioscience (Extended Data Fig. 1c). Probes with high predicted likelihoods of off-target pull-down (for example, such as those in high-repeat regions) were masked and removed from the probe tiling, and probe coverage was double-checked to ensure the inclusion of key genomic features (for example, all promoters and CTCF sites in the locus) before finalization. Probe panels were synthesized and purchased as Custom Target Enrichment Panels from Twist Bioscience.

Capture of target loci. Capture was performed in accordance with Twist Bioscience's Standard Hybridization Target Enrichment Protocol. Briefly, pooled sample libraries were dried and mixed with Hybridization Mix (Twist Bioscience #104178), Custom Panels (Twist Bioscience #101001) and Universal Blockers (Twist Bioscience #100578), as well as Mouse Cot-1 DNA (Invitrogen #18440016). The library pool was hybridized to the biotinylated probe panel overnight, after which streptavidin beads (Twist Bioscience #100983) were used to pull down probes with hybridized ligated fragments and then washed (Twist Bioscience #104178) to remove unbound fragments. Another round of PCR amplified the target-enriched library using the Equinox Library Amplification Mix (Twist Bioscience #104178), including a test PCR (as described above) to identify the number of amplification cycles necessary to meet sequencing requirements. With 2–4 µg input library for capture, the RCMC samples generated in this paper needed five to six cycles of post-capture PCR amplification. Following PCR amplification, the captured library was purified (Twist Bioscience #100983) and then quantified via both Fragment Analyzer and qPCR at the MIT BioMicro Center in preparation for sequencing submission.

Three technical replicates of the pre-capture Micro-C library were generated for each of the four initially tested conditions (WT, 45 min transcriptional inhibition with triptolide, RAD21 depletion and a RAD21 depletion control), after which each replicate was simultaneously captured for the *Klf1*, *Ppm1g* and *Fbn2* loci. After the publication of TMCC[17], additional probes for the *Sox2* and *Nanog* loci were designed and a single additional capture experiment was conducted pooling all three pre-capture Micro-C libraries for simultaneous *Sox2* and *Nanog* capture. Subsequently, a biological replicate was generated for each of the initially tested conditions, with the inclusion of the additional perturbation of 4 h transcriptional inhibition with triptolide. Pre-capture Micro-C libraries were constructed for each of the five conditions, after which each library was simultaneously captured for all five target loci. Finally, a biological replicate of the 4 h transcriptional inhibition perturbation was generated; once a pre-capture Micro-C library for it was generated, it was pooled with the WT library from the first technical replicate of the first biological replicate, and the pooled libraries were simultaneously captured for all five target loci.

Sequencing. Following qPCR quantification, post-capture libraries across samples (WT, transcriptionally inhibited, cohesin depleted and DMSO-treated control) were pooled in a 1:1 molar ratio. Pooled libraries were sequenced by paired-end 2 × 50 cycle sequencing kits with Illumina NovaSeq SP or S1 flow cells on a NovaSeq 6000 system by the Broad Institute of MIT and Harvard's Walk-Up Sequencing services. Basecalls for NovaSeq output were performed using bcl2fastq v2.20.0.422.

Data analysis

Mapping and normalizing RCMC. RCMC paired-end reads generated by the Illumina NovaSeq sequencers were downloaded as .fastq files for each sample, pair mate, and flow cell lane. Read quality was verified using FastQC (v0.11.9). Paired end reads were aligned to the UCSC mm39 genome using bowtie2 (v2.3.5.1) with −local −reorder −very-sensitive-local. Aligned paired-end reads were then parsed with pairtools (v0.3.0) parse with −add-columns mapq −walks-policy mask −min-mapq 2. Parsed reads were filtered for PCR duplicates and unmapped/multiple mapping reads with pairtools dedup with −max-mismatch 1. Remaining reads were indexed (pairix v0.3.7) and filtered (pairtools select) to retain only those reads where both read mates lay in a locus of interest. These filtered reads were subsequently converted to .cool format using cooler (v0.8.11) cload pairs, creating binned read counts across the genome for 50-bp bins. Finally, .cool files were converted to the .mcool format with cooler zoomify including the −balance option, compiling read counts for bins from 50 bp up to 10 Mb in size.

Contact matrices were balanced using iterative correction and eigendecomposition (ICE)[30], which normalizes all rows and columns of a contact matrix sum to the same value. Applying ICE balancing to all mapped reads generated subpar normalization and generated an artifact where 'stripes' containing no capture probe coverage appeared to have greater contact densities than adjacent probe-covered regions (Extended Data Fig. 1e). ICE balancing to .mcool files containing data only within captured regions of interest (ROIs) did not result in these artifacts, and was therefore used in for all RCMC data in this study. The success of ICE balancing applied to these ROI-only .mcools was evaluated against published whole-genome Hi-C[31] and Micro-C[12] datasets in mESCs (Extended Data Fig. 6a). The sum of each row of each of the RCMC, Micro-C and Hi-C balanced contact matrices at 250-bp resolution within capture ROIs was calculated, plotted as a histogram distribution of row sums and verified to match the distribution of column sums. The subset of RCMC rows containing microcompartment anchors was also plotted to confirm that they match the distribution of row sums across the whole locus, ruling out that microcompartments are an artifact of incomplete ICE normalization[30].

Visualizing RCMC. RCMC contact maps were visualized alongside genomic annotations, published ChIP-seq, RNA-seq and ATAC-seq datasets using the HiGlass[54] browser (http://higlass.io/) and software (v0.8.0). Contact maps shown in figures were generated using cooltools (v0.5.0) (https://cooltools.readthedocs.io/). Genomic tracks (that is, ChIP-seq, RNA-seq and ATAC-seq) and gene annotations in figures were generated using CoolBox[55] (v0.3.3). In generating our genomic tracks, we analyzed 27 public datasets (Supplementary Table 1) using processed bigWig files that were CrossMapped[56] (v0.6.1) (http://crossmap.sourceforge.net/) to the mm39 reference genome. Tracks were visualized using the Integrative Genomics Viewer[57] (v2.10.3) to scale tracks by identifying local maxima and minimizing noise.

All other bioinformatic and data analyses are provided as Supplementary Methods.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data generated in this study can be found at NCBI Gene Expression Omnibus under accession number GSE207225. Source data are provided with this paper.

Code availability

All custom code and scripts used for data analyses in this paper are available on GitHub at https://github.com/ahansenlab/RCMC_analysis_code and on Zenodo at https://zenodo.org/record/7641852. Source data are provided with this paper.

References

52. Pettitt, S. J. et al. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat. Methods* **6**, 493–495 (2009).
53. Hansen, A. S. et al. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6**, e25776 (2017).
54. Kerpedjiev, P. et al. HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol.* **19**, 125 (2018).
55. Xu, W. et al. CoolBox: a flexible toolkit for visual analysis of genomics data. *BMC Bioinformatics* **22**, 489 (2021).
56. Zhao, H. et al. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics* **30**, 1006–1007 (2014).
57. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
58. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res.* **27**, 1939–1949 (2017).
59. Venev, S. et al. open2c/cooltools: v0.4.1 (v0.4.1). *Zenodo* https://doi.org/10.5281/zenodo.5214125 (2021).
60. Abdennur, N. et al. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
61. Robinson, J. T. et al. Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst.* **6**, 256–258 (2018).

Author contributions

V.Y.G. and A.S.H. designed the project. V.Y.G. developed and optimized the RCMC protocol with input from M.K.H. and A.S.H. V.Y.G. performed cell culture and RCMC, analyzed sequencing data, compared RCMC to previous studies, and performed loop and compartment calling, identification of individual microcompartments and validation of ICE balancing. M.K.H. performed and analyzed western blotting and

a Example MNase Titration

b Example Gel Extraction

c Probe Design Workflow

1. Identify ROI(s) of interest

2. Design a tiling of 80-mer biotinylated dsDNA probes across ROI(s)

3. Remove probes targeting highly repetitive regions that cause off-target capture

4. Order / synthesize probes *(Twist Bioscience)*

d Capture efficiency by locus

RCMC Locus	*Ppm1g*	*Klf1*	*Fbn2*	*Sox2*	*Nanog*
Chromosome	chr5	chr8	chr18	chr3	chr6
Coordinates *(mm39)*	31,257,344–32,382,344	84,846,629–85,856,629	58,032,072–59,034,072	33,804,149–35,704,149	122,451,959–122,876,959
Locus Size (Mb)	1.125	1.010	1.002	1.900	0.425

		Ppm1g	*Klf1*	*Fbn2*	*Sox2*	*Nanog*
Biological Rep 1	Mapped Reads *(separate Capture sets)*		777,589,080		184,490,998	
	Mapped Reads *(to locus)*	101,312,622	246,236,115	31,555,452	41,871,251	22,726,278
	Capture Efficiency		49%		35%	
Biological Rep 2	Mapped Reads *(combined Capture)*		383,565,523			
	Mapped Reads *(to locus)*	38,296,105	83,552,946	11,700,380	36,566,842	18,325,332
	Capture Efficiency		49%			

e Contact map normalization by ICE

Klf1 **Locus** (chr8, 1kb res)

ICE balanced, data in Capture loci only

Unbalanced data

ICE balanced, data in Capture loci only

ICE balanced, all sequenced data

85.6 Mb   85.7 Mb   85.8 Mb   *mm39*

f RCMC matches Hi-C & Micro-C at the *Fbn2* TAD (chr18, 2kb res)

RCMC

RCMC

Hi-C (Bonev)

Micro-C (Hsieh)

58.1 Mb   58.3 Mb   58.5 Mb   *mm39*

CTCF
SMC1A

Fbn2

Fbn2

g RCMC has high technical and biological reproducibility

Ppm1g Locus (chr5)

NA	0.98	0.99	0.99	BR2_WT
0.98	NA	1.00	1.00	TR1_WT
0.99	1.00	NA	1.00	TR2_WT
0.99	1.00	1.00	NA	TR3_WT

BR2_WT TR1_WT TR2_WT TR3_WT

Klf1 Locus (chr8)

NA	0.98	0.98	0.99	BR2_WT
0.98	NA	1.00	1.00	TR1_WT
0.98	1.00	NA	1.00	TR2_WT
0.99	1.00	1.00	NA	TR3_WT

BR2_WT TR1_WT TR2_WT TR3_WT

Fbn2 Locus (chr18)

NA	1.00	0.99	0.99	TR2_WT
1.00	NA	0.99	1.00	TR3_WT
0.99	0.99	NA	0.99	BR2_WT
0.99	1.00	0.99	NA	TR1_WT

TR2_WT TR3_WT BR2_WT TR1_WT

Sox2 Locus (chr3)

NA	0.76	0.78	0.78	TR3_WT
0.76	NA	0.98	0.98	BR2_WT
0.78	0.98	NA	0.99	TR1_WT
0.78	0.98	0.99	NA	TR2_WT

TR3_WT BR2_WT TR1_WT TR2_WT

Nanog Locus (chr6)

NA	0.94	0.95	0.95	TR3_WT
0.94	NA	0.99	0.99	BR2_WT
0.95	0.99	NA	1.00	TR1_WT
0.95	0.99	1.00	NA	TR2_WT

TR3_WT BR2_WT TR1_WT TR2_WT
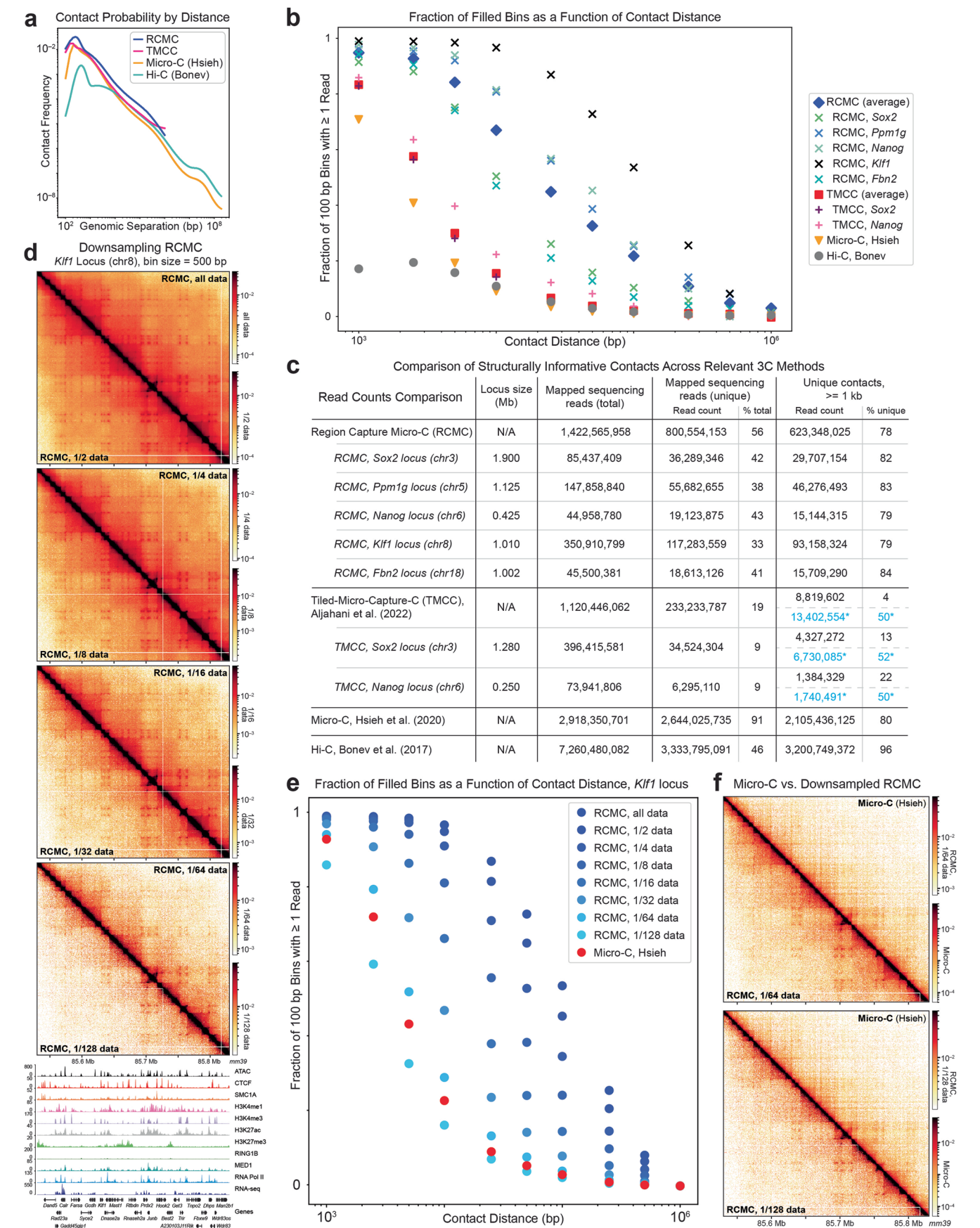
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | RCMC efficiently and reproducibly captures ligated dinucleosomal fragments, giving rise to deep contact maps. (**a**) Representative MNase titration DNA gel indicating the ideal level of digestion by MNase, based on the ratio of fragment sizes, for the RCMC protocol. (**b**) Representative size-selection gel for the RCMC protocol showing the dinucleosomal band that is extracted to obtain ligated fragments. (**c**) Overview of the capture probe design workflow for RCMC. 80-mer probes tiling the region of interest are designed, removing those which overlap highly repetitive regions. (**d**) Summary of the capture efficiency for each of the five regions for which probes were designed. The locations and sizes of the regions, the number of ligated fragments which mapped at single loci at both ends in total and in the region, and the capture efficiencies are given. Because different capture probe sets were used for Biological Replicates 1 (two separate sets of capture probes) and 2 (simultaneous capture for all five loci), numbers are separately provided for each Biological Replicate. (**e**) Contact maps comparing raw,

unbalanced data (upper panel, lower triangle), ICE-balanced[30] to all aligned reads (lower panel, lower triangle) and ICE-balanced to reads in captured loci only (both panels, upper triangle). Balancing only to data entirely within captured loci was necessary to remove artifacts due to capture bias. (**f**) Contact maps comparing the entire *Fbn2* TAD in RCMC and in Hi-C[31] and Micro-C[12]. Gene annotations and ChIP-seq signal tracks are shown below the contact maps. (**g**) Measurement of reproducibility between WT replicates across all five capture loci, with reproducibility scores determined using HiCRep[58] at 10 kb resolution, clustered according to similarity. Three technical RCMC replicates (denoted by 'TR#') comprise Biological Replicate 1, while 'BR2' denotes Biological Replicate 2. TR3_WT is noted in blue text at the *Sox2* and *Nanog* loci because very little TR3_WT pre-Capture library remained for input to *Sox2* & *Nanog* capture after the initial *Ppm1g*, *Klf1* and *Fbn2* capture experiment; accordingly, relative to all other replicates, TR3_WT has much lower sequencing depth (0.5–2.4% the number of unique contacts) at the *Sox2* & *Nanog* loci.

a Contact Probability by Distance

b Fraction of Filled Bins as a Function of Contact Distance

d Downsampling RCMC
Klf1 Locus (chr8), bin size = 500 bp

c Comparison of Structurally Informative Contacts Across Relevant 3C Methods

Read Counts Comparison	Locus size (Mb)	Mapped sequencing reads (total)	Mapped sequencing reads (unique)		Unique contacts, >= 1 kb	
			Read count	% total	Read count	% unique
Region Capture Micro-C (RCMC)	N/A	1,422,565,958	800,554,153	56	623,348,025	78
RCMC, Sox2 locus (chr3)	1.900	85,437,409	36,289,346	42	29,707,154	82
RCMC, Ppm1g locus (chr5)	1.125	147,858,840	55,682,655	38	46,276,493	83
RCMC, Nanog locus (chr6)	0.425	44,958,780	19,123,875	43	15,144,315	79
RCMC, Klf1 locus (chr8)	1.010	350,910,799	117,283,559	33	93,158,324	79
RCMC, Fbn2 locus (chr18)	1.002	45,500,381	18,613,126	41	15,709,290	84
Tiled-Micro-Capture-C (TMCC), Aljahani et al. (2022)	N/A	1,120,446,062	233,233,787	19	8,819,602 / 13,402,554*	4 / 50*
TMCC, Sox2 locus (chr3)	1.280	396,415,581	34,524,304	9	4,327,272 / 6,730,085*	13 / 52*
TMCC, Nanog locus (chr6)	0.250	73,941,806	6,295,110	9	1,384,329 / 1,740,491*	22 / 50*
Micro-C, Hsieh et al. (2020)	N/A	2,918,350,701	2,644,025,735	91	2,105,436,125	80
Hi-C, Bonev et al. (2017)	N/A	7,260,480,082	3,333,795,091	46	3,200,749,372	96

e Fraction of Filled Bins as a Function of Contact Distance, *Klf1* locus

f Micro-C vs. Downsampled RCMC

Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Benchmarking of RCMC against other 3C methods.
(**a**) Contact probability curves comparing RCMC against the highest resolution
Tiled-Micro-Capture-C (TMCC)[17], Micro-C[12], and Hi-C[31] mESC datasets across
contact distances. (**b**) Benchmarking comparison of RCMC's ability to fill
out high-resolution contact matrices against TMCC[17], Micro-C[12], and Hi-C[31].
Region-averaged calculations are shown for all methods, and calculations for
individual captured regions are also shown for RCMC and TMCC. The x axis
shows the contact distance in bp, and the y axis shows the fraction of all bins at a
given contact distance within the captured locus that contain at least one read at
100 bp resolution. (**c**) Summary of read counts across RCMC, TMCC[17], Micro-C[12],
and Hi-C[31]. The number of mapped sequencing reads, the fraction of unique
reads, and the fraction of structurally informative (defined as cis contacts > =1 kb)
unique reads are given for each method. Two versions of quantification
are provided for TMCC. In black are numbers processed using the same
bioinformatic pipeline as for RCMC. Capture region-specific quantifications

(defined here as all reads with at least one of two read mates mapped to the locus)
are also provided for all RCMC loci and the $Sox2$ and $Nanog$ TMCC loci; the $Oct4$
and $Prdm14$ TMCC loci are not considered in this manuscript. In blue are numbers
kindly provided by Dr. A Marieke Oudelaar, obtained using the custom TMCC-
specific bioinformatic pipeline from Aljahani et al.[17]. Values with asterisks denote
quantifications of all unique contact pairs mapped to captured loci (not filtered
to be > = 1 kb in size). (**d**) Contact map comparisons of RCMC data generated
in this manuscript, starting from the full dataset (topmost) and successively
downsampled by orders of two down to 1/128th of the data (bottommost), shown
for the $Klf1$ locus at 500 bp resolution. (**e**) As in (b), benchmarking comparison
of successively downsampled RCMC's ability to fill out high-resolution contact
matrices against Micro-C[12] at the $Klf1$ locus. (**f**) Contact map comparisons of
1/64th and 1/128th downsampled RCMC (left) against the highest-resolution
available mESC Micro-C[12] (right; Hsieh 2020) dataset, shown for the $Klf1$ locus
at 500 bp resolution.
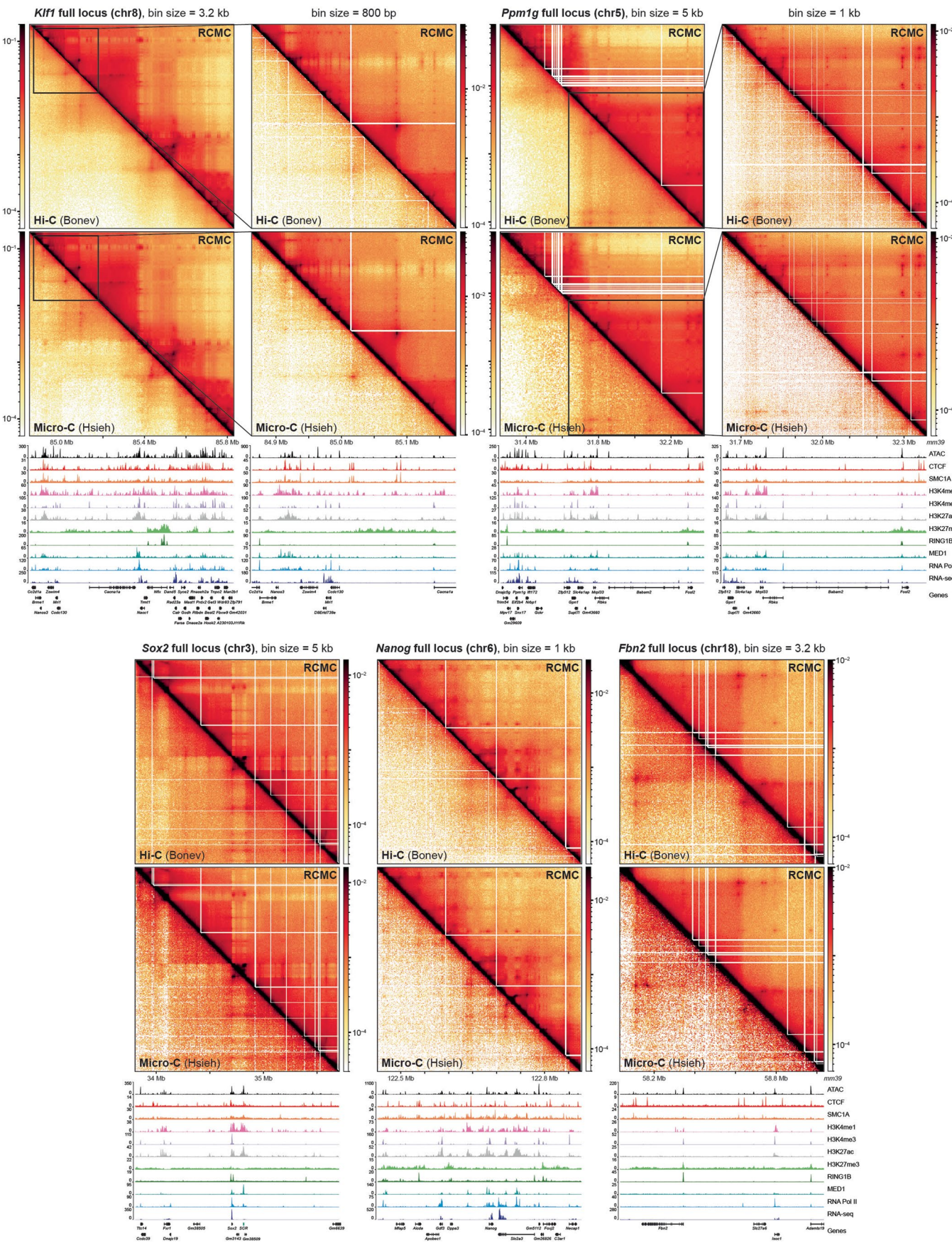
**Extended Data Fig. 3 | See next page for caption.**

Extended Data Fig. 3 | RCMC generates deeper contact maps than other 3C methods across all 5 captured loci. Contact map comparisons of RCMC against the highest-resolution available mESC Hi-C[31] (top; Bonev et al. 2017) and Micro-C[12] (bottom; Hsieh et al. 2020) datasets at the *Klf1*, *Ppm1g*, *Sox2*, *Nanog*, and *Fbn2* loci. Full captured regions are shown for each locus at resolutions ranging from 1–5 kb, as well as *Klf1* and *Ppm1g* zoom-ins at 800 and 1000 bp, respectively. Gene annotations and ATAC, ChIP-seq, and RNA-seq tracks (Supplementary Table 1) are shown below the contact maps, while the contact intensity scales are shown next to the maps.
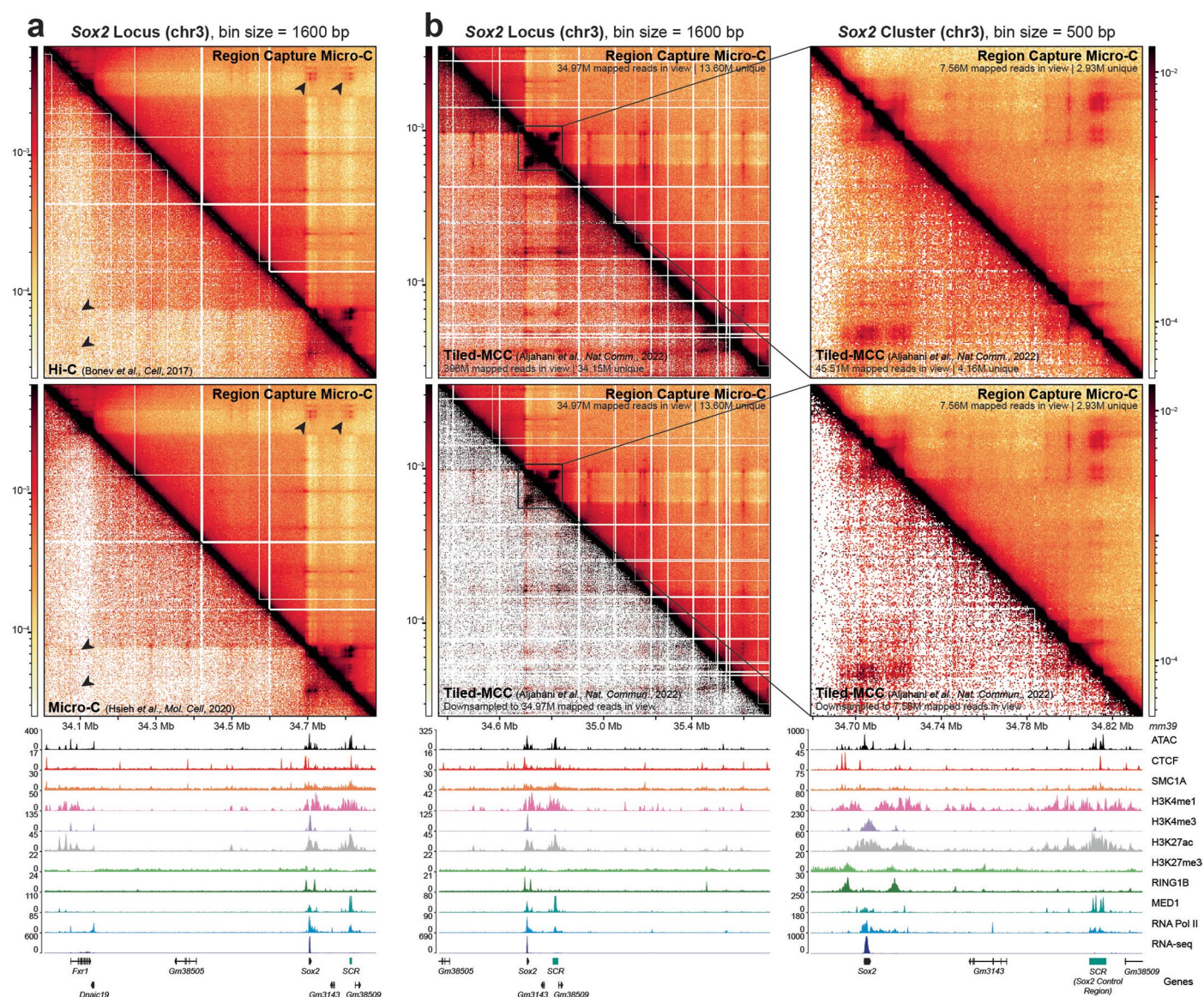
Extended Data Fig. 4 | RCMC maps the *Sox2* locus more deeply and efficiently than sister methods, uncovering previously unseen interactions. (a) Contact map comparisons of RCMC against Hi-C[31] (top) and Micro-C[12] (bottom) at the *Sox2* locus at 1.6 kb resolution. Arrows mark contacts between *Sox2*, the SCR, and *Fxr1* not mapped by Hi-C and Micro-C. **(b)** Contact map comparisons of RCMC against Tiled-Micro-Capture-C[17] (TMCC) across the whole TMCC-Captured locus (left, 1600 bp resolution) and in the *Sox2* and SCR regulatory cluster (right, 500 bp resolution). Full datasets are visualized in the top contact maps, and TMCC has been downsampled to match the total number of RCMC sequencing reads in view in the bottom contact maps.
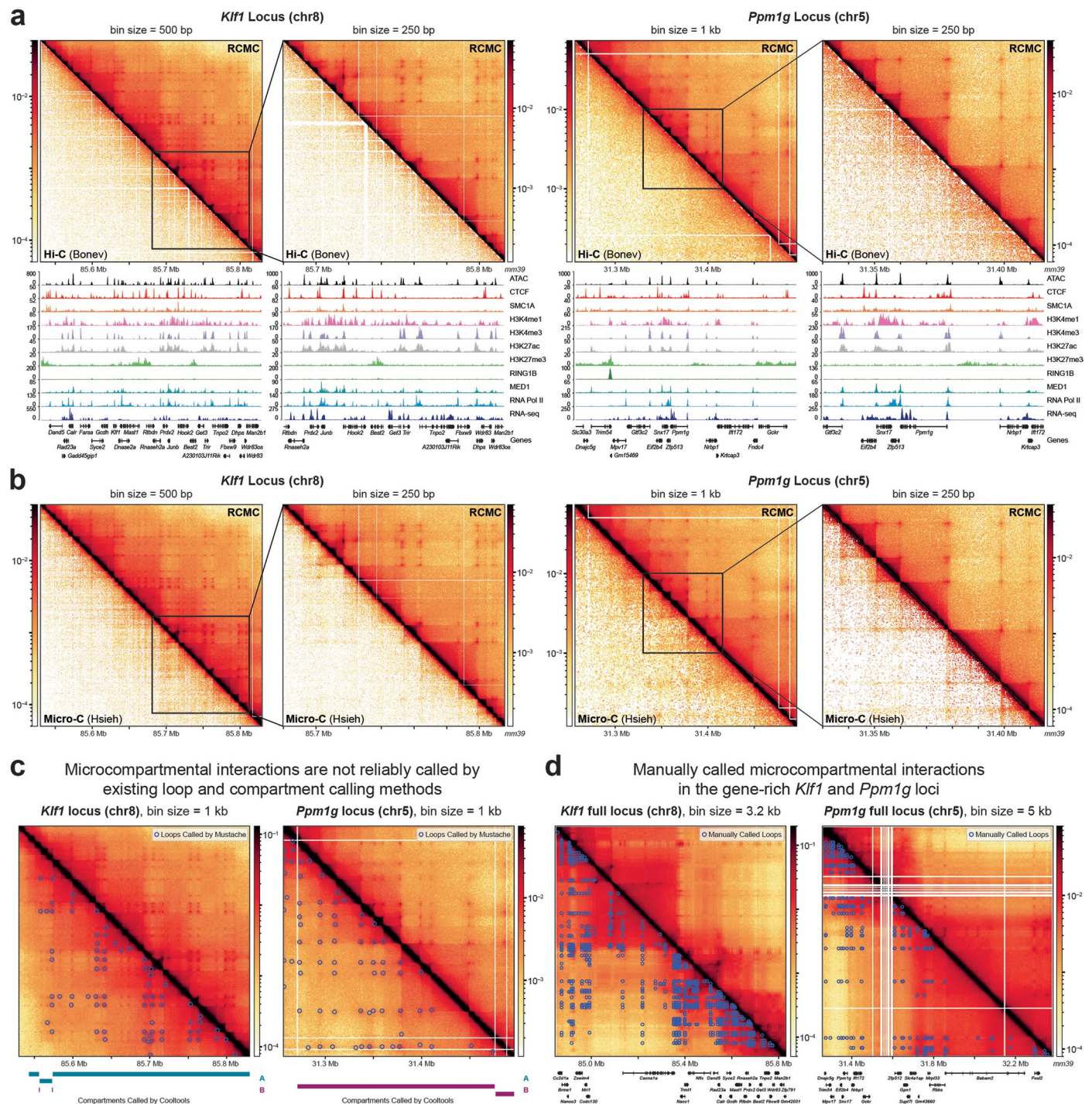
Extended Data Fig. 5 | RCMC identifies microcompartments, which are not visible in other methods and not reliably called by existing algorithms. (a, b) Contact maps comparison of RCMC (top) against Hi-C[31] (bottom, **a**) and Micro-C[12] (bottom, **b**) at the *Klf1* locus at 500 and 250 bp resolutions and at the *Ppm1g* locus at 1000 and 250 bp resolutions. (**c**) Contact maps of the *Klf1* and *Ppm1g* loci at 1 kb resolution with loop calls by Mustache[37] overlaid on the bottom half of the map and compartment calls by cooltools[59,60] shown below the map. (**d**) Contact maps of the entire *Klf1* (3.2 kb resolution) and *Ppm1g* (5 kb resolution) captured loci with manually called loops (see Methods) overlaid on the bottom halves of the maps.

a Microcompartments are not the result of incomplete ICE balancing
Matrix rowsums across methods and for microcompartment anchors are similar

b Microcompartment anchors can be ATAC-negative, and not all ATAC peaks are anchors

c Contact maps reveal non-ATAC microcompartment anchors
and ATAC peaks that do not form microcompartmental interactions

d Fewer than half of all ATAC peaks in the *Klf1* and *Ppm1g* loci form microcompartments

Extended Data Fig. 6 | Microcompartments are not artifacts resulting from incomplete ICE balancing nor chromatin accessibility. (**a**) Comparison of ICE balancing across methods and captured loci. Distributions of the sums of ICE-balanced contact matrix rows at 250 bp resolution are shown at the *Klf1*, *Ppm1g*, *Fbn2*, and *Sox2* loci for RCMC, Micro-C[12], and Hi-C[31], as well as for the subset of RCMC rows containing microcompartment anchors. A sharp unimodal peak is consistent with ICE's baseline assumption that all contact matrix rows and columns must sum to the same value. (**b**) Metaplots (above) and heatmaps (below) depicting ATAC signal at microcompartment anchors (left, separated by whether anchors coincide with an ATAC peak) and at all ATAC peaks in the *Klf1* and *Ppm1g* capture loci (right, separated by whether peaks coincide with a microcompartment anchor). Signals are plotted in a 2 kb window centered on the anchor (left) or the ATAC peak (right). (**c**) RCMC contact maps at the *Klf1*

(left, 250 bp resolution) and *Ppm1g* (right, 1.6 kb resolution) loci indicating ATAC peaks that do not form microcompartments (left, magenta arrows) and a microcompartment anchor that does not coincide with an ATAC peak (right, cyan arrow). Black arrows (right) indicate microcompartmental loops involving the ATAC-negative microcompartment anchor. (**d**) Venn diagram breakdown of the overlap between all manually annotated microcompartment anchors and all ATAC peaks across the *Klf1* and *Ppm1g* capture loci. Of 132 annotated microcompartment anchors, 12 do not coincide with ATAC peaks (cyan) while 120 do (purple, *). Of 353 called ATAC peaks, 187 do not form microcompartment anchors (magenta) while 166 do (purple, **). The apparent discrepancy of 120 microcompartment anchors being anchored by 166 ATAC peaks is due to two close ATAC peaks occasionally anchoring a single microcompartment.
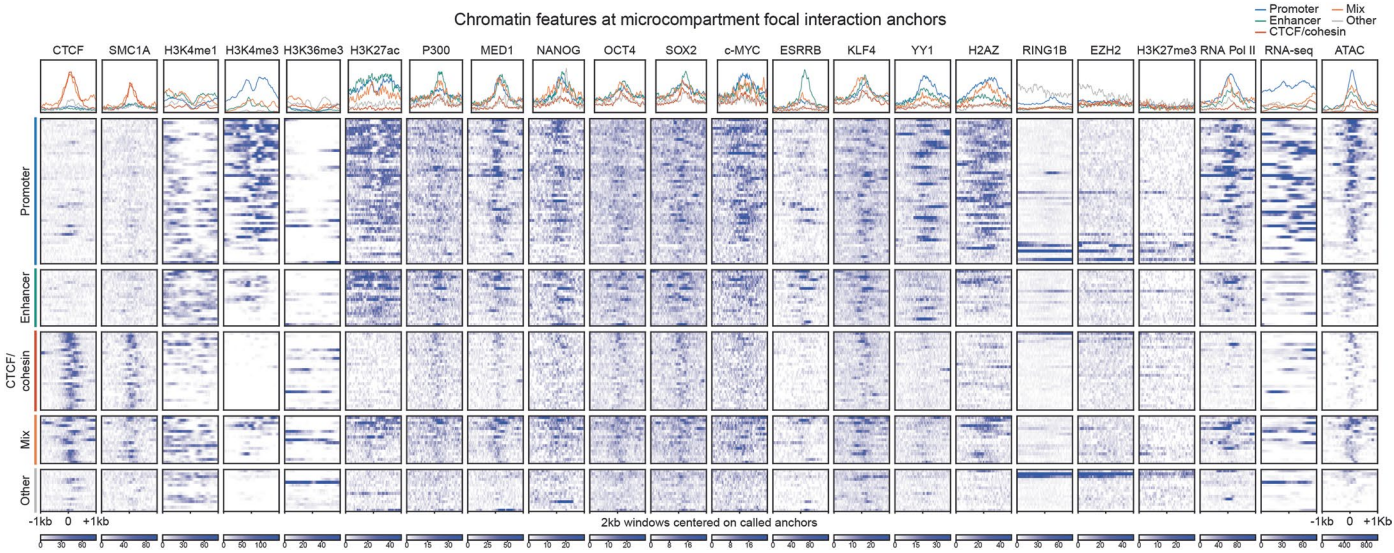
Extended Data Fig. 7 | Categories of microcompartment anchors can be defined by their chromatin features. Metaplots (above) and heatmaps (below) depicting ATAC, ChIP-seq, and RNA-seq (Supplementary Table 1) signal at microcompartment loop anchors for classes of microcompartment anchors as defined in Fig. 3e. Features are plotted in a 2 kb window centered on the anchor.
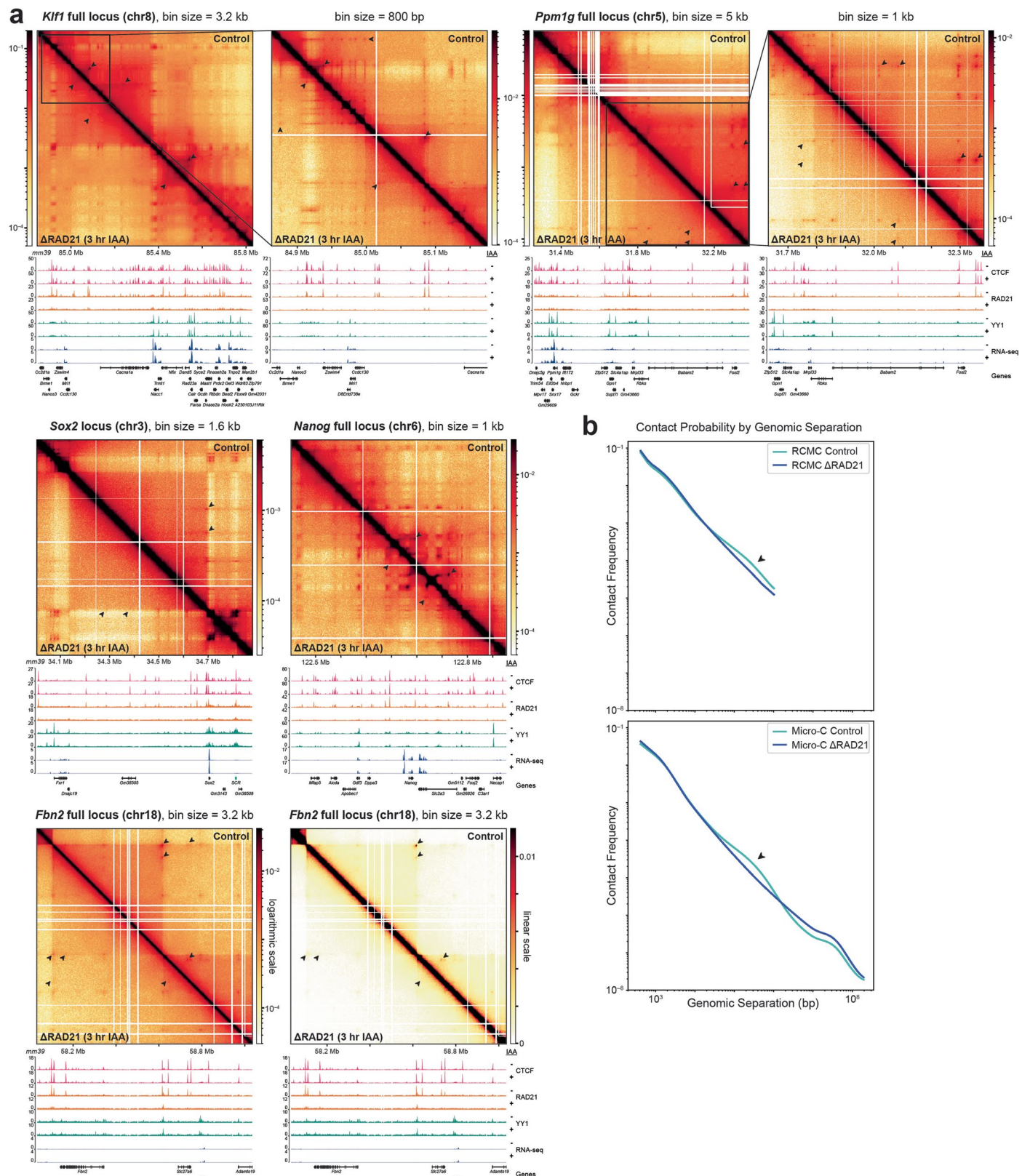
Extended Data Fig. 8 | Cohesin depletion disrupts CTCF/Cohesin loops, but generally not most microcompartmental loops. (**a**) Contact maps comparing a DMSO control (above) and RAD21-depleted samples (below) are shown for the *Klf1*, *Ppm1g*, *Sox2*, *Nanog*, and *Fbn2* loci at resolutions spanning 800 bp – 5 kb in F1M RAD21-mAID-BFP-V5 mESCs[13,38]. Arrows mark contacts lost upon RAD21 depletion. ChIP-seq data from Hsieh et al.,[13] is shown below the maps before and after the IAA treatment (500 μM, 3 hours). Two versions of the *Fbn2* locus are shown, with the left using logarithmic contact frequency scaling and the right using linear scaling. Loss of the *Fbn2* loop[38] is most clearly seen on linear scale. (**b**) Contact probability curves comparing RAD21-depleted RCMC samples against a DMSO control (top) and RAD21-depleted Micro-C samples against a DMSO control (bottom). Arrows indicate the contact frequency 'bump' lost upon RAD21 depletion.
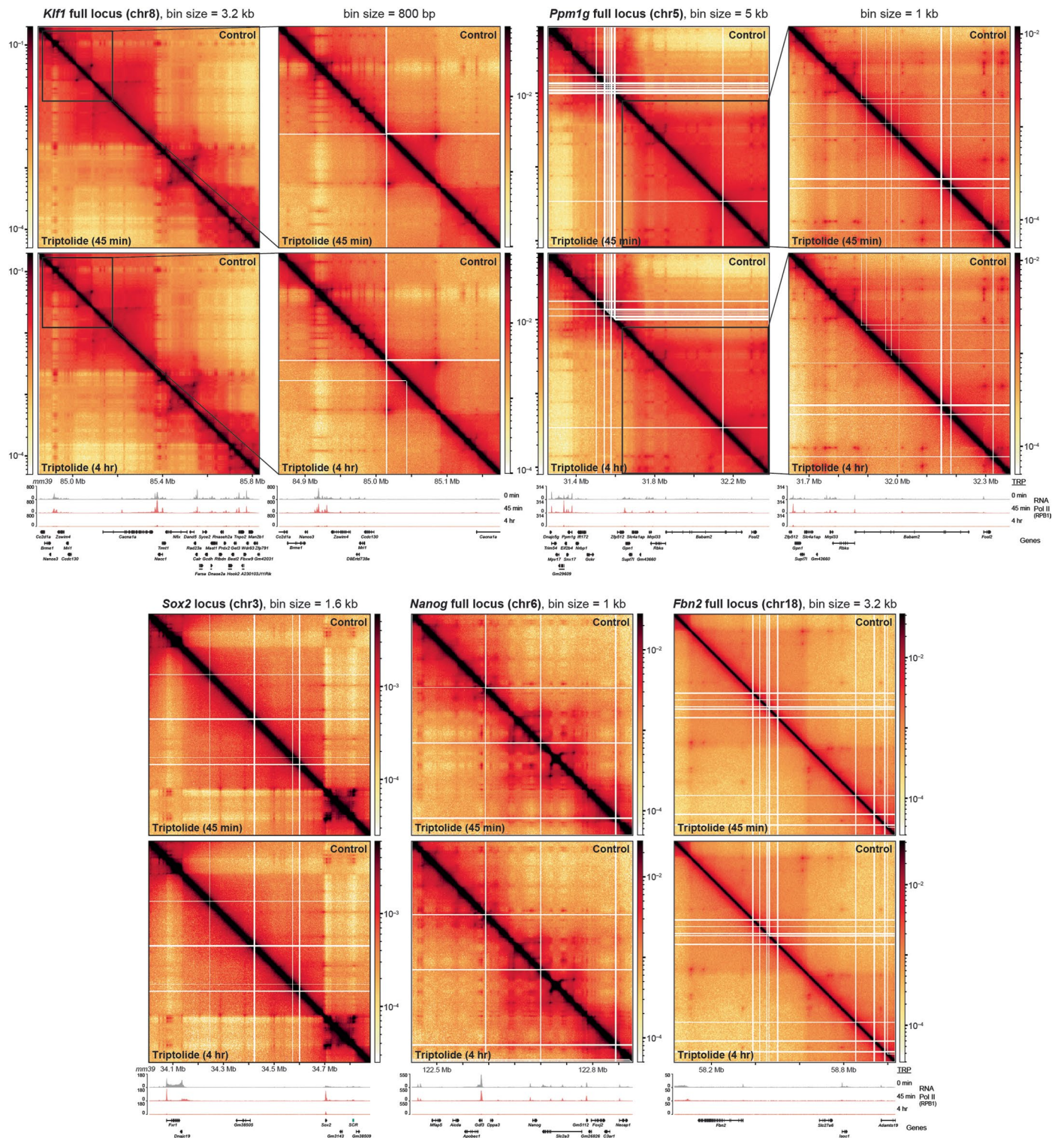
Extended Data Fig. 9 | Inhibition of transcription does not significantly alter genome organization in captured loci. Contact maps comparing control data against 45 min (top) and 4 hr (bottom) transcriptional inhibition data (from 1 μM triptolide treatments) are shown for the *Klf1, Ppm1g, Sox2, Nanog,* and *Fbn2* loci at resolutions spanning 800 bp – 5 kb in mESC WT cells. RNA Pol II ChIP-seq data is shown below the maps for each treatment condition.
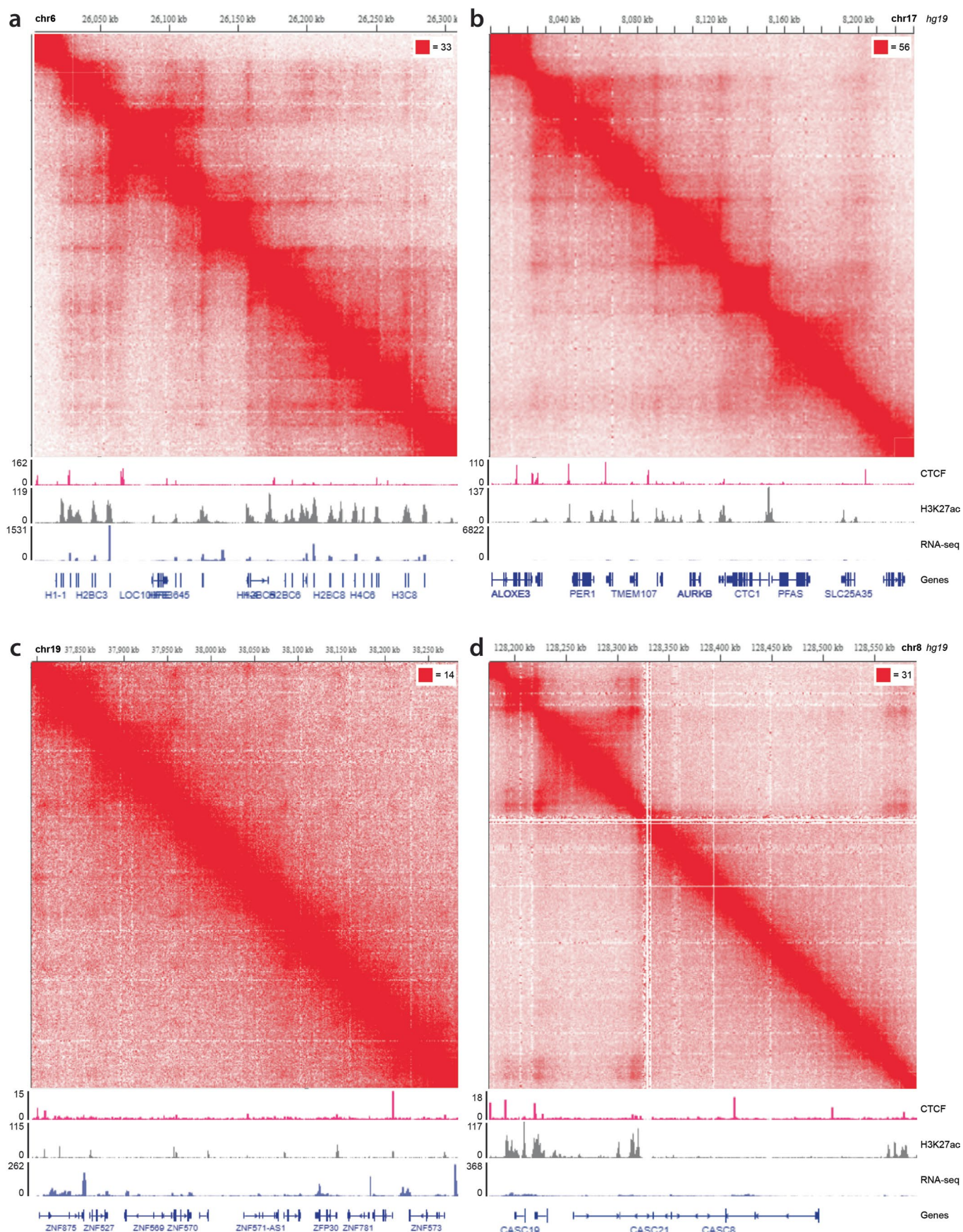
Extended Data Fig. 10 | Microcompartment-like structures are also visible in ultra-deep Hi-C data. (a–d) Contact maps of ultra-deep Hi-C data in human lymphoblastoid cells Harris et al.,[14] showing loci with structures sharing many microcompartmental features. Maps were generated using Juicebox's web interface[61] kindly provided by Dr. Jordan Rowley. Maps are shown at 1 kb resolution, with GM12878 gene annotations, CTCF (ENCFF364OXN) and H3K27ac (ENCFF180LKW) ChIP-seq, and RNA-seq (ENCFF604VIC) signal tracks shown below the contact maps.

Corresponding author(s): Anders S. Hansen

Last updated by author(s): Mar 3, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	bcl2fastq v2.20.0.422, fastqc v0.11.9, bowtie2 v2.3.5.1, pairtools v0.3.0, pairix v0.3.7, cooler v0.8.11, higlass v0.8.0, cooltools v0.5.0, coolbox v0.3.3, crossmap v0.6.1, IGV v2.10.3, hicrep v1.12.2, mustache v1.2.4, chromosight v1.6.1, sip v1.6.1, bigWigToBedGraph v377, MACS2 v2.2.7.1, FIMO v5.4.1, bedtools v2.30.0, R v4.1.2, GenomicRanges v1.46.1, ggplot2 v3.3.6, deeptools v3.5.1, Python 3.7.12

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data generated in this study can be found at NCBI GEO under accession number GSE207225 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207225. A list of publicly available datasets used in this study is provided in Supplementary Table 1. All custom code and scripts used for data analyses in

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Not applicable
Population characteristics	Not applicable
Recruitment	Not applicable
Ethics oversight	Not applicable

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes were determined based on the sample sizes chosen for similar experiments and analyses using other chromosome conformation capture techniques. Specifically, prior Micro-C (e.g., Hsieh et al., (2020)) and Capture (e.g., Oudelaar et al., (2020)) studies reflected the cellular inputs required to achieve sufficient library complexity for fine-scale genome architecture mapping as well as the sequencing depth necessary to resolve structures.
Data exclusions	No data were excluded.
Replication	Experimental findings reported in this study were reliably reproduced in multiple technical and biological replicates, and replicate-to-replicate reproducibility was quantitatively confirmed using HiCRep. The first biological cell sample was fixed in July 2021, from which three technical replicates were generated between October-December 2021 for each of the tested conditions. The second biological replicate was fixed July 2022, from which a single technical replicate was generated for each of the tested conditions; an additional treatment of 4 hr transcriptional inhibition was tested in this RCMC replicate. Subsequently, a second biological replicate of the 4 hr transcriptional inhibition was generated in October 2022.
Randomization	Randomization was not relevant to this study as samples were determined by the known genotypes of cells used and the treatments applied to them. Therefore, randomization was not appropriate nor relevant.
Blinding	Blinding was not relevant to this study. The RCMC method had not previously been developed and validated; as a result, there was no prior knowledge of results for any of the tested conditions.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Rabbit polyclonal anti-RAD21, Abcam, Cat# ab154769, lot GR3224138-16 Mouse monoclonal anti-TBP, Abcam, Cat# ab818, clone number 1TBP18 Rabbit monoclonal anti-RPB1 NTD, Cell Signaling Technology, Cat# 14958 HRP-linked Sheep anti-Mouse IgG whole Ab, Cytiva Life Sciences, Cat# NA931 HRP-linked Donkey anti-Rabbit IgG whole Ab, Cytiva Life Sciences, Cat# NA934
Validation	anti-RAD21 – manufacturer-validated in K562, THP1, HL60, NIH3T3, JC, BCL-1, and NCI-H929 cells by western blot, validated here by protein depletion in mouse cell line anti-TBP – manufacturer-validated in U2OS and HeLa cells by ChIP-qPCR and cellular fractionation anti-RPB1 NTD – manufacturer-validated in HeLa, KNRK, and COS7 cells by western blot, and in HeLa cells by ChIP-qPCR and ChIP-seq HRP-linked Sheep anti-Mouse IgG – manufacturer-validated using affinity adsorption to remove cross-reacting antibodies to other species followed by mouse IgG affinity column purification HRP-linked Donkey anti-Rabbit IgG – manufacturer-validated using affinity adsorption to remove cross-reacting antibodies to other species followed by rabbit IgG affinity column purification

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Wild-type JM8.N4 mouse embryonic stem cells (mESCs), an established cell line (Research Resource Identifier: RRID:CVCL_J962; obtained from the KOMP Repository at UC Davis), were used in this study. F1M RAD21-mAID-BFP-V5 JM8.N4 mESCs were previously generated and validated in the Hansen lab (Gabriele et al. Science 2022). HEK293T cells were obtained from ATCC (CRL-3216). No new cell lines were generated in this study.
Authentication	All cell lines utilized in this study have previously been validated by PCR, sequencing, and western blotting, and by fluorescence microscopy where appropriate.
Mycoplasma contamination	All cell lines were regularly tested for mycoplasma contamination and were found to be negative. mESC cell lines were further pathogen tested using an IMPACT II test and found negative for all tested pathogens.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	The data generated in this study can be found at NCBI GEO under accession number GSE207225 at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE207225.
Files in database submission	RCMCp_allCap_WT_mm39.merged.50.mcool RCMCp_allCap_TI_mm39.merged.50.mcool RCMCp_allCap_DMSO_mm39.merged.50.mcool RCMCp_allCap_IAA_mm39.merged.50.mcool TMC_R1_211104_TI_ME_R1.fastq.gz TMC_R1_211104_DMSO_ME_R1.fastq.gz TMC_R1_211104_IAA_ME_R1.fastq.gz TMC_R1_211104_TI_ME_R2.fastq.gz TMC_R1_211104_DMSO_ME_R2.fastq.gz TMC_R1_211104_IAA_ME_R2.fastq.gz TMC_R2_211031_WT_ME_R1.fastq.gz TMC_R2_211031_TI_ME_R1.fastq.gz TMC_R2_211031_DMSO_ME_R1.fastq.gz TMC_R2_211031_IAA_ME_R1.fastq.gz TMC_R2_211031_WT_ME_R2.fastq.gz TMC_R2_211031_TI_ME_R2.fastq.gz TMC_R2_211031_DMSO_ME_R2.fastq.gz TMC_R2_211031_IAA_ME_R2.fastq.gz TMC_R2_211119_WT_ME_R1.fastq.gz TMC_R2_211119_TI_ME_R1.fastq.gz TMC_R2_211119_DMSO_ME_R1.fastq.gz TMC_R2_211119_IAA_ME_R1.fastq.gz TMC_R2_211119_WT_ME_R2.fastq.gz TMC_R2_211119_TI_ME_R2.fastq.gz TMC_R2_211119_DMSO_ME_R2.fastq.gz TMC_R2_211119_IAA_ME_R2.fastq.gz TMC_R3_211119_WT_ME_R1.fastq.gz

```
TMC_R3_211119_TI_ME_R1.fastq.gz
TMC_R3_211119_DMSO_ME_R1.fastq.gz
TMC_R3_211119_IAA_ME_R1.fastq.gz
TMC_R3_211119_WT_ME_R2.fastq.gz
TMC_R3_211119_TI_ME_R2.fastq.gz
TMC_R3_211119_DMSO_ME_R2.fastq.gz
TMC_R3_211119_IAA_ME_R2.fastq.gz
TMC_SN_211219_WT_R1_ME_P1.fastq.gz
TMC_SN_211219_TI_R1_ME_P1.fastq.gz
TMC_SN_211219_DMSO_R1_ME_P1.fastq.gz
TMC_SN_211219_IAA_R1_ME_P1.fastq.gz
TMC_SN_211219_WT_R1_ME_P2.fastq.gz
TMC_SN_211219_TI_R1_ME_P2.fastq.gz
TMC_SN_211219_DMSO_R1_ME_P2.fastq.gz
TMC_SN_211219_IAA_R1_ME_P2.fastq.gz
TMC_SN_211219_WT_R2_ME_P1.fastq.gz
TMC_SN_211219_TI_R2_ME_P1.fastq.gz
TMC_SN_211219_DMSO_R2_ME_P1.fastq.gz
TMC_SN_211219_IAA_R2_ME_P1.fastq.gz
TMC_SN_211219_WT_R2_ME_P2.fastq.gz
TMC_SN_211219_TI_R2_ME_P2.fastq.gz
TMC_SN_211219_DMSO_R2_ME_P2.fastq.gz
TMC_SN_211219_IAA_R2_ME_P2.fastq.gz
TMC_SN_211219_WT_R3_ME_P1.fastq.gz
TMC_SN_211219_TI_R3_ME_P1.fastq.gz
TMC_SN_211219_DMSO_R3_ME_P1.fastq.gz
TMC_SN_211219_IAA_R3_ME_P1.fastq.gz
TMC_SN_211219_WT_R3_ME_P2.fastq.gz
TMC_SN_211219_TI_R3_ME_P2.fastq.gz
TMC_SN_211219_DMSO_R3_ME_P2.fastq.gz
TMC_SN_211219_IAA_R3_ME_P2.fastq.gz
RCMC_BR1_merged_allCap_WT_mm39.merged.50.mcool
RCMC_BR1_merged_allCap_TI_mm39.merged.50.mcool
RCMC_BR1_merged_allCap_DMSO_mm39.merged.50.mcool
RCMC_BR1_merged_allCap_IAA_mm39.merged.50.mcool
RCMC_BR1_LT_mm39_allCap.merged.50.mcool
RCMC_BR2_WT_mm39_allCap.merged.50.mcool
RCMC_BR2_TI_mm39_allCap.merged.50.mcool
RCMC_BR2_DMSO_mm39_allCap.merged.50.mcool
RCMC_BR2_IAA_mm39_allCap.merged.50.mcool
RCMC_BR2_LT_mm39_allCap.merged.50.mcool
WT_xInput_r2_mm39_MERGED_rmdup_downsampled.bw
Trp45min_xInput_r2_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_xInput_r1_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_xInput_r2_mm39_MERGED_rmdup_downsampled.bw
WT_PolII_xChIP_r2_mm39_MERGED_rmdup_downsampled.bw
Trp45min_PolII_xChIP_r2_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_PolII_xChIP_r1_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_PolII_xChIP_r2_mm39_MERGED_rmdup_downsampled.bw
ME_HKVFTDRX2.RCMC_R1_WT.1.fastq.gz
ME_HKVFTDRX2.RCMC_R1_WT.2.fastq.gz
RCMC_BR1_LT_mm39_allCap.merged.50.mcool
RCMC_BR2_WT_mm39_allCap.merged.50.mcool
RCMC_BR2_TI_mm39_allCap.merged.50.mcool
RCMC_BR2_DMSO_mm39_allCap.merged.50.mcool
RCMC_BR2_IAA_mm39_allCap.merged.50.mcool
RCMC_BR2_LT_mm39_allCap.merged.50.mcool
WT_xInput_r2_mm39_MERGED_rmdup_downsampled.bw
Trp45min_xInput_r2_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_xInput_r1_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_xInput_r2_mm39_MERGED_rmdup_downsampled.bw
WT_PolII_xChIP_r2_mm39_MERGED_rmdup_downsampled.bw
Trp45min_PolII_xChIP_r2_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_PolII_xChIP_r1_mm39_MERGED_rmdup_downsampled.bw
Trp4hr_PolII_xChIP_r2_mm39_MERGED_rmdup_downsampled.bw
RCMC_BR2_220904_LT_ME_R1.fastq.gz
RCMC_BR2_220904_WT_ME_R1.fastq.gz
RCMC_BR2_220904_TI_ME_R1.fastq.gz
RCMC_BR2_220904_DMSO_ME_R1.fastq.gz
RCMC_BR2_220904_IAA_ME_R1.fastq.gz
ME_HKVFTDRX2.RCMC_RepLT.1.fastq.gz
ME_HKVFTDRX2.WT_xInput_r2.1.fastq.gz
ME_HKVFTDRX2.Trp45min_xInput_r2.1.fastq.gz
ME_HKVFTDRX2.Trp4h_xInput_r1.1.fastq.gz
ME_HKVFTDRX2.Trp4h_xInput_r2.1.fastq.gz
ME_HKVFTDRX2.WT_RNAPII_xChIP_r2.1.fastq.gz
```

ME_HKVFTDRX2.Trp45min_RNAPII_xChIP_r2.1.fastq.gz
ME_HKVFTDRX2.Trp4h_RNAPII_xChIP_r1.1.fastq.gz
ME_HKVFTDRX2.Trp4h_RNAPII_xChIP_r2.1.fastq.gz
RCMC_BR2_220904_LT_ME_R2.fastq.gz
RCMC_BR2_220904_WT_ME_R2.fastq.gz
RCMC_BR2_220904_TI_ME_R2.fastq.gz
RCMC_BR2_220904_DMSO_ME_R2.fastq.gz
RCMC_BR2_220904_IAA_ME_R2.fastq.gz
ME_HKVFTDRX2.RCMC_RepLT.2.fastq.gz
ME_HKVFTDRX2.WT_xInput_r2.2.fastq.gz
ME_HKVFTDRX2.Trp45min_xInput_r2.2.fastq.gz
ME_HKVFTDRX2.Trp4h_xInput_r1.2.fastq.gz
ME_HKVFTDRX2.Trp4h_xInput_r2.2.fastq.gz
ME_HKVFTDRX2.WT_RNAPII_xChIP_r2.2.fastq.gz
ME_HKVFTDRX2.Trp45min_RNAPII_xChIP_r2.2.fastq.gz
ME_HKVFTDRX2.Trp4h_RNAPII_xChIP_r1.2.fastq.gz
ME_HKVFTDRX2.Trp4h_RNAPII_xChIP_r2.2.fastq.gz

Genome browser session
(e.g. UCSC)

No longer applicable

Methodology

Replicates

ChIP-seq was performed as a control for triptolide treatment alongside RCMC experiments. ChIP-seq was performed for the second WT and 45-minute triptolide treatment biological replicates, and both replicates of 4 hour triptolide treatments. Metaplot analysis of RNA PolII ChIP-seq across gene bodies in 4 hour treatment indicated strong agreement between replicates.

Sequencing depth

WT_Input_rep2: total reads – 9951099, uniquely mapped reads – 7223157, read length – 63 bp, paired end
Trp45min_Input_rep2: total reads – 14122377, uniquely mapped reads – 10212908, read length – 63 bp, paired end
Trp4hr_Input_rep1: total reads – 13513860, uniquely mapped reads – 9703548, read length – 63 bp, paired end
Trp4hr_Input_rep2: total reads – 2080839, uniquely mapped reads – 1502470, read length – 63 bp, paired end
WT_PolIIChIP_rep2: total reads – 117139446, uniquely mapped reads – 87802126, read length – 63 bp, paired end
Trp45min_PolIIChIP_rep2: total reads – 66899880, uniquely mapped reads – 49291256, read length – 63 bp, paired end
Trp4hr_PolIIChIP_rep1: total reads – 60345970, uniquely mapped reads – 43340740, read length – 63 bp, paired end
Trp4hr_PolIIChIP_rep2: total reads – 61915886, uniquely mapped reads – 45177491, read length – 63 bp, paired end

Antibodies

Rabbit monoclonal anti-Rpb1-NTD, Cell Signaling Technology, Cat# 14958

Peak calling parameters

Peaks were not called using data generated in this study. Peaks using data from previous studies (CTCF, GSE90994; SMC1A, GSE123636; H3K27ac, GSE90893; H3K4me1, ENCFF282RLA) were called using MACS2 bdgpeakcall on bedgraph files generated from bigwigs available on GEO for each dataset.

Data quality

Quality of ChIP-seq data for RNA PolII was assessed by inspection of individual replicate bigwig files and by comparison to published datasets of the same treatment in the same cell line, and by comparison of treatment condition data by metaplot and heatmap, and correlation analysis, using deepTools.

Software

Paired end reads were aligned to concatenated mouse (mm39) and human (hg38) genomes using Bowtie2 using this script: spikeinChIP_PE_alignment.py, https://zenodo.org/record/7641852#.ZAIpax_MJD8). bigwig files were generated using MACS2 pileup function and converted from wig format using wigToBigWig from UCSC. Metaplot and heatmap analyses of ChIP-seq data were performed using computeMatrix and plotProfile/plotHeatmap in deepTools.