ARTICLE

doi:10.1038/nature24281

Two independent modes of chromatin organization revealed by cohesin removal

Wibke Schwarzer[1]*, Nezar Abdennur[2]*, Anton Goloborodko[3]*, Aleksandra Pekowska[4], Geoffrey Fudenberg[5], Yann Loe-Mie[6,7], Nuno A Fonseca[8], Wolfgang Huber[4], Christian H. Haering[9], Leonid Mirny[3,5] & Francois Spitz[1,4,6,7]

Imaging and chromosome conformation capture studies have revealed several layers of chromosome organization, including segregation into megabase-sized active and inactive compartments, and partitioning into sub-megabase domains (TADs). It remains unclear, however, how these layers of organization form, interact with one another and influence genome function. Here we show that deletion of the cohesin-loading factor Nipbl in mouse liver leads to a marked reorganization of chromosomal folding. TADs and associated Hi-C peaks vanish globally, even in the absence of transcriptional changes. By contrast, compartmental segregation is preserved and even reinforced. Strikingly, the disappearance of TADs unmasks a finer compartment structure that accurately reflects the underlying epigenetic landscape. These observations demonstrate that the three-dimensional organization of the genome results from the interplay of two independent mechanisms: cohesin-independent segregation of the genome into fine-scale compartments, defined by chromatin state; and cohesin-dependent formation of TADs, possibly by loop extrusion, which helps to guide distant enhancers to their target genes.

The three-dimensional organization of chromosomes is tightly related to their biological function[1–3]. Genome-wide chromosome conformation capture (Hi-C) maps have revealed key features of the 3D organization of metazoan chromosomes, including compartmentalization, TADs, and interaction peaks[4,5] (Extended Data Fig. 1). Compartmentalization is visible as a characteristic checkerboard pattern of contact enrichment both in *cis* and in *trans* between megabase-sized genomic intervals of the same type, reflecting spatial segregation of transcriptionally active (type A) and inactive (type B) chromatin[6]. TADs appear as squares of enriched contact frequency with sharp boundaries; they usually span hundreds of kilobases and do not necessarily exhibit any checkering[7,8]. TADs are thought to contribute to gene expression, notably by promoting or preventing interactions between promoters and distant regulatory elements[9–11]. Finally, peaks (often termed loops[12]) are visible as focal enrichments in contact frequency between pairs of loci, often at the corners of TAD squares. Despite a lack of supporting mechanistic experiments, compartments, TADs and peaks are typically assumed to constitute hierarchical levels of chromosome folding; however, the connection between them remains poorly understood.

Architectural proteins, notably cohesin and the transcriptional insulator CTCF, are believed to have crucial roles in chromatin organization during interphase[13]. Cohesin and CTCF co-localize at TAD boundaries and the bases of Hi-C peaks[7,12], but their roles have not been fully defined. The recently proposed loop extrusion model[14,15] yields predictions that are consistent with previous experimental results. In this model, TADs emerge from the progressive extrusion of chromatin loops by a protein complex (for example, cohesin) until it dissociates from chromatin or reaches a boundary element (for example, CTCF sites). Recent experimental manipulation of CTCF motifs[15–17] and CTCF depletion[18] support the proposed role of CTCFs as boundary elements, but the role of cohesin in interphase chromatin organization and the process of loop extrusion remains unclear, as experimental depletions of cohesin have shown limited impact on chromatin organization[19–21].

To identify the role of cohesin in interphase chromatin, we deleted *Nipbl* (yeast homologue *Scc2*), which is necessary for loading of cohesin onto chromatin[22]. As the turnover period of chromosome-bound cohesin is about 20 minutes[23,24], constant loading is required for its presence on DNA. We achieved efficient deletion of *Nipbl* in non-dividing hepatocytes by using a liver-specific, tamoxifen-inducible Cre driver (Fig. 1a), which circumvents the lethality of $Nipbl^{+/-}$ mice and the essentiality of cohesin in dividing cells (Extended Data Fig. 2). Ten days after tamoxifen injection, *Nipbl* expression was greatly reduced (Fig. 1b), leading to displacement of cohesin from the chromatin fraction to the soluble nuclear fraction, which indicates loss of cohesin from chromosomes (Fig. 1c). This strong depletion of chromatin-bound cohesin was also observed by calibrated chromatin immunoprecipitation and sequencing (ChIP–seq)[25] for RAD21 and SMC3 both genome-wide (Extended Data Fig. 3) and at wild-type cohesin peaks (Fig. 1d, e). A conservative estimate indicates at least a fourfold to sixfold decrease in chromatin-associated cohesin. The liver showed no particular pathological signs compared to control mice (mock-injected $Nipbl^{flox/flox}$; *Ttr-cre/Esr1* mice or tamoxifen-injected $Nipbl^{+/+}$; *Ttr-cre/Esr1* mice). Hepatocytes showed no sign of cell death or proliferation (Extended Data Fig. 2).

To assess the consequences of *Nipbl* depletion and cohesin loss on chromosome organization, we performed tethered chromatin conformation capture[26] (a variant of Hi-C) on purified hepatocytes from wildtype, tamoxifen control (TAM) and Δ*Nipbl* mice (Fig. 1f). For each of these three conditions, two biological replicates were generated; five out of six replicates produced more than 25 million

[1]Developmental Biology Unit. European Molecular Biology Laboratory. 69117 Heidelberg, Germany. [2]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts USA. [3]Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts USA. [4]Genome Biology Unit. European Molecular Biology Laboratory. 69117 Heidelberg, Germany. [5]Institute for Medical Engineering and Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts USA. [6]Institut Pasteur, (Epi)genomics of Animal Development Unit, Developmental and Stem Cell Biology Department. Institut Pasteur. 75015 Paris, France. [7]CNRS, UMR3738, 25 rue du Dr Roux, 75015 Paris, France. [8]European Bioinformatics Institute. European Molecular Biology Laboratory. Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, UK. [9]Cell Biology and Biophysics Unit, European Molecular Biology Laboratory, 69117 Heidelberg, Germany.
*These authors contributed equally to this work.

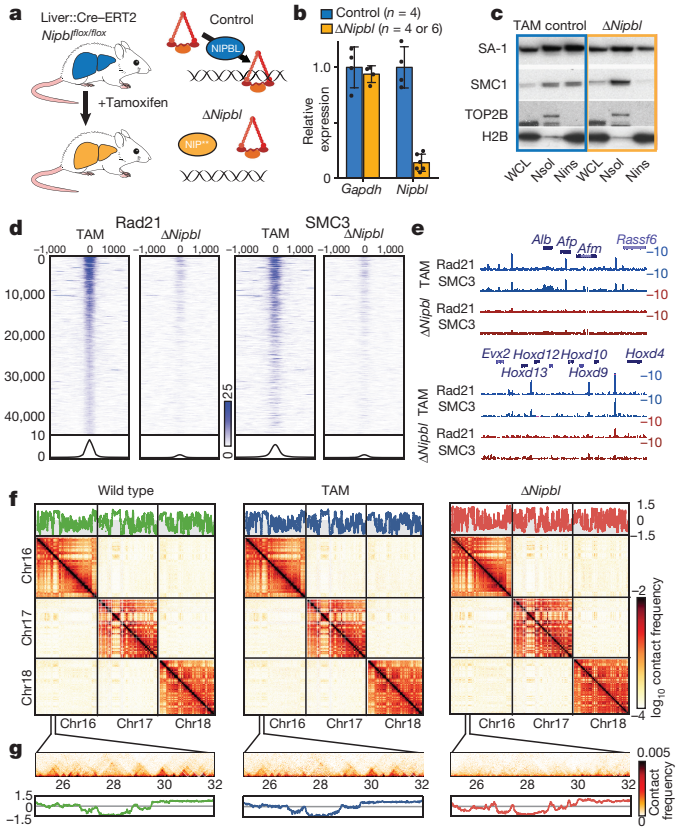


Figure 1 | Overview of the experimental design. a, We deleted a conditional allele of *Nipbl* (Extended Data Fig. 2) in adult hepatocytes using a liver-specific driver for the Cre–ERT2 fusion protein after injection of tamoxifen. In the absence of *Nipbl,* cohesin (red triangular ring) is not loaded on chromatin. **b**, Expression of *Nipbl* and *Gapdh* (control) by quantitative PCR with reverse transcription (RT–qPCR) in control ($n = 4$) and Δ*Nipbl* ($n = 6$ for *Nipbl*; 4 for *Gapdh*) hepatocytes. Mean normalized gene expression (using *Pgk1* as internal control, and wild-type expression set as 1) is displayed as mean ± s.d. and compared using unpaired two-sided *t*-test (95% confidence interval (CI): control = (0.7033–1.297); Δ*Nipbl* = (0.05731–0.2198) for *Nipbl* expression). **c**, Western blots of hepatocyte protein extracts (WCL, whole cell lysate; Nsol, nuclear extract, soluble fraction; Nins, insoluble, chromatin fraction) showed displacement of cohesin structural subunits (SA-1 and SMC1) from the chromatin-bound fraction. Histone H2B and topoisomerase 2β (TOP2B) distributions serve as controls for loading and enrichment of two nuclear fractions. Experiment repeated on two biologically independent samples per condition. See Supplementary Data 1 for gel source data. **d**, Stacked heatmaps of calibrated chromatin immunoprecipitation and sequencing (ChIP–seq) signal (for Rad21 and SMC3) at wild-type Rad21 peaks ranked by fold change over input, in the TAM control condition. **e**, ChIP–seq tracks for Rad21 and SMC3 over representative genomic regions. **f**, **g**, Hi-C maps at 20 kb resolution of wild-type (left), TAM control (middle) and Δ*Nipbl* cells (right). Top, *cis* compartment tracks; middle, *cis* and *trans* contact maps of chromosomes 16–18. **g**, An example of a short-range Hi-C map at chr 16:25–32 Mb with compartment tracks.

interactions at separations of over 10 kb, on par with other primary tissue Hi-C results[27] (Supplementary Tables 1, 2). The contact maps obtained from each biological replicate showed extensive similarities (Supplementary Data 2), allowing us to pool the two replicate data sets to generate Hi-C maps for the three different conditions. We compared Hi-C maps for Δ*Nipbl* and control cells by examining compartments, TADs, peaks, and global scaling of the contact probability $P(s)$[28] (Extended Data Fig. 1).

Disappearance of TADs and peaks

Our Hi-C maps reveal that *Nipbl* deletion has a striking effect on genome organization (Fig. 1f, g), contrasting with the very mild changes

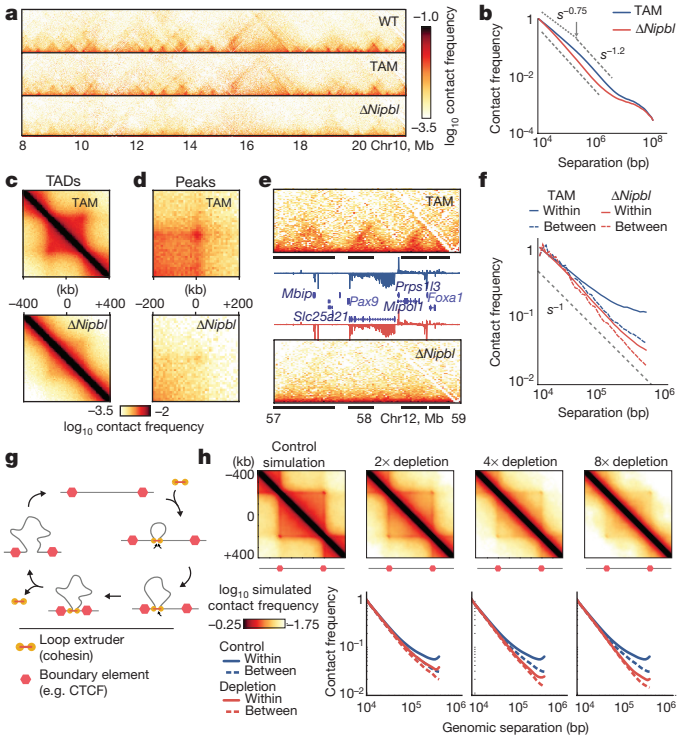


Figure 2 | *Nipbl* deletion leads to disappearance of TADs and peaks from Hi-C contact maps. a, Hi-C map for chr10:8–21 Mb illustrating loss of TADs and peaks. **b**, Genome-wide $P(s)$ curves for TAM control and Δ*Nipbl* cells, normalized to unity at $s = 10$ kb. **c**, **d**, Mean Hi-C map of 564 TADs of length 300–400 kb (**c**) and 102 peaks of separation 500–600 kb (**d**). **e**, Hi-C map of an example 2-Mb region chr12:57–59 Mb (top, TAM control; bottom, Δ*Nipbl*) with expression tracks (middle, annotated genes and RPM-normalized RNA-seq; sense above axis, antisense below; TAM in blue, Δ*Nipbl* in red). Black bars show wild-type TADs. Hi-C and RNA-seq experiments were repeated independently on two and four biological replicates, respectively, with similar results. **f**, $P(s)$ curves plotted separately for contacts formed within or between TADs of size 300–500 kb. **g**, Cartoon representation of the loop extrusion model of cohesin action[14]. In this model, cohesins form *cis*-loops by first binding to adjacent loci on chromosomes (top and right). After binding, cohesins translocate along the fibre in both directions, effectively extruding a loop (bottom). Extrusion halts when cohesins reach boundary elements. Extruded loops disassemble when cohesins unbind from the chromosome (left). **h**, Polymer simulations of loop extrusion reproduce the effects of cohesin depletion. Top row, simulated contact maps, showing average TADs. Left to right, the impact of sequential cohesin depletion on the contact map of a TAD in simulations. Bottom row, $P(s)$ calculated separately within and between TADs.

reported in previous cohesin depletion experiments[19–21] (Extended Data Fig. 4). Compared to wild-type and TAM control samples, Δ*Nipbl* cells show genome-wide disappearance of local TAD patterns (Fig. 1g) but persistence of type A–type B compartmentalization (Fig. 1f). Disappearance of TADs in Δ*Nipbl* cells is widespread and can be seen in individual maps (Fig. 2a), as well as on the composite map constructed by averaging the Δ*Nipbl* Hi-C map around locations of domain boundaries detected in wild-type maps (Fig. 2c, Extended Data Fig. 5). Some local organization is retained in regions with higher activity (A compartments), where cohesin is about threefold more abundant than in inactive regions (B compartments) in TAM cells (Extended Data Figs 5, 2e). We show below that these structures are not residual or new TADs, but unmasked small compartments. TAD-associated peaks of contact enrichment also disappear in Δ*Nipbl* maps, showing up to fourfold reduction in contact frequency (Fig. 2d and Extended Data Fig. 4a), notably between convergent CTCF sites. Insulation and directionality of the contact footprint of CTCF sites are also absent in

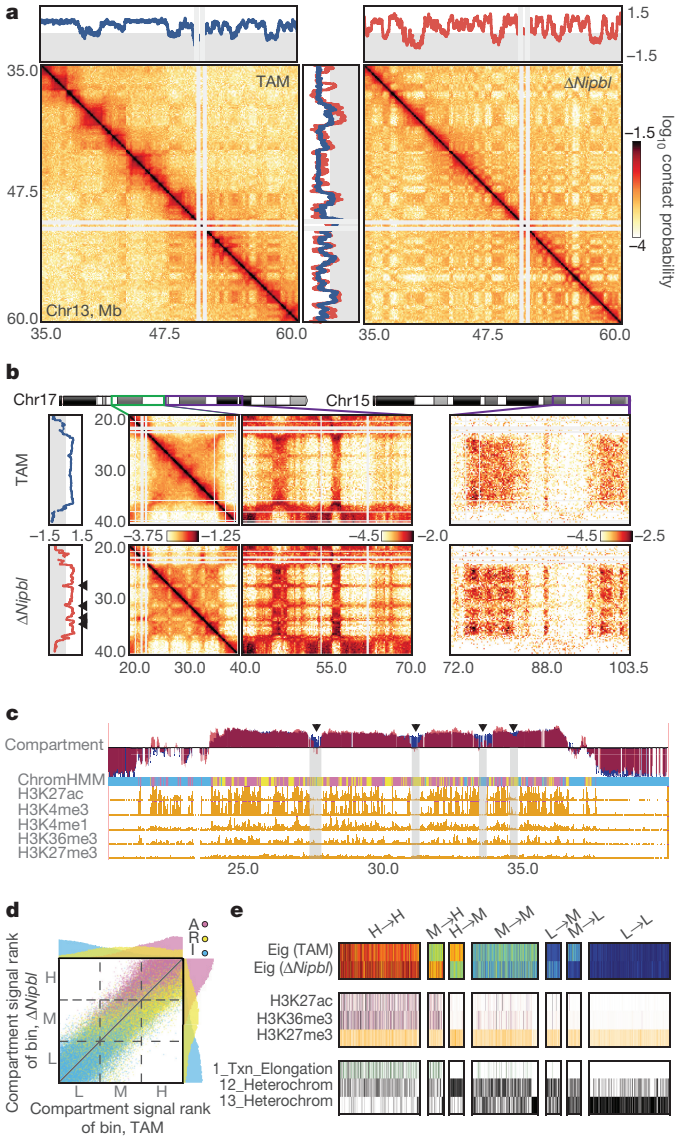


Figure 3 | *Nipbl* deletion leads to activity-dependent alteration of compartment structure. a, An example region (chr13: 35–60 Mb) showing changes in compartmentalization. Top and centre, 20-kb *cis* compartment tracks. **b**, Hi-C maps of an example region at 200 kb with predominantly positive (type A) compartment signal in TAM control (top) showing fragmentation upon *Nipbl* deletion (bottom), manifested in the alternating contact patterns of short-range (<10 Mb, middle left) and long-range *cis* (middle right) and *trans* contact maps (right). In **a** and **b**, the two Hi-C replicates of each condition show similar results. **c**, Loci that experience a local drop in the compartment signal are depleted in epigenetic marks of activity. Top to bottom: compartment track in TAM (blue) and Δ*Nipbl* (red); simplified ChromHMM state assignments (magenta, active; yellow, repressed; cyan, inert); ENCODE activity-related histone ChIP–seq for adult mouse liver cells. In **b** and **c**, arrowheads indicate local drops in compartment signal. **d**, Rank correlation of 20-kb compartment tracks ($n = 113{,}372$ 20-kb genomic bins) in TAM and Δ*Nipbl* cells, coloured by simplified ChromHMM state. Top and right margins, histograms of compartment signal ranks split by simplified ChromHMM state in Δ*Nipbl* (right) and TAM (top). The dashed lines show the tercile borders, splitting bins into equal-sized groups of low (L), middle (M) and high (H) compartment signal. **e**, Epigenetic profiles of bins transitioning between compartment signal terciles upon *Nibpl* deletion. Top to bottom: compartment track in wild-type and Δ*Nipbl* cells; ENCODE histone marks; ChromHMM states characteristic of active, repressed and inert chromatin. The bins that transitioned from the middle to the high tercile are enriched in activity marks, whereas bins that transitioned from the high to the middle tercile were depleted in those marks.

Δ*Nipbl* cells (Extended Data Fig. 6). However, we observe no effect of *Nipbl* deletion on CTCF occupancy, demonstrating that the loss of TADs and the CTCF contact footprint in Δ*Nipbl* cells is not due to loss of or changes in CTCF occupancy (Extended Data Fig. 3). This finding strengthens the idea that CTCF and cohesin have distinct roles in shaping chromosome architecture[14,18].

The changes seen in Δ*Nipbl* cells cannot be attributed to altered gene expression, since TAD patterns vanished equally in regions where gene expression was unchanged (Fig. 2e) and in those where it was upregulated or downregulated (Extended Data Fig. 5c). This major reorganization of chromatin architecture is also reflected in plots of contact frequency $P(s)$ as a function of genomic distance s[3,29]. In control samples, as in other mammalian cells[14,15], the $P(s)$ curve has two regimes: a more shallow decay for $s < 200$ kb ($P(s) \approx s^{-0.7}$), and a steeper scaling for 200 kb $< s <$ 3 Mb ($P(s) \approx s^{-1.2}$). Loss of chromatin-associated cohesin led to disappearance of the first regime, producing a single decay of contact probability across the whole range (Fig. 2b). This observation suggests that the first scaling regime reflects the compaction of the genome associated with TADs. We confirmed this by calculating $P(s)$ separately within and between wild-type TAD intervals. In control cells, $P(s)$ within TADs decreases more slowly than $P(s)$ between TADs. In Δ*Nipbl* cells, the within-TAD $P(s)$ curve collapses to the between-TAD $P(s)$ curve, indicating that the characteristic enrichment of contacts within TADs is lost and that chromatin folding becomes more uniform and decompacted (Fig. 2f), consistent with decompaction observed by imaging upon *Nipbl* reduction[30].

Next, we simulated the effects of *Nipbl* depletion in our model of loop extrusion, by reducing the number of extruding cohesins (Fig. 2g). For each simulated concentration of extruding cohesins, we calculated Hi-C maps and $P(s)$ within and between TADs. In these simulations, eightfold depletion of extruding cohesins was required to achieve agreement with our experimental data, manifested by (i) noticeable disappearance of TAD and corner peak enrichments; (ii) loss of the $P(s) \approx s^{-0.7}$ regime in the scaling; and (iii) decompaction of chromatin (Fig. 2h). Together, these analyses and the observed effects of *Nipbl* deletion indicate that cohesin has a central role in the local compaction of chromosomes, and support the idea that this effect is mediated by the production of dynamic populations of extruded chromatin loops between boundary elements, which form TAD and corner peak patterns in interphase Hi-C maps[14,15].

Enhanced and finer compartmentalization

We next examined the compartmentalization of chromatin in Δ*Nipbl* cells. As noted earlier, in contrast to the marked loss of TADs, compartmentalization still exists (Fig. 1) and is enhanced about 1.8-fold (Extended Data Fig. 5d, e; Supplementary Data 2, 3). Moreover, closer examination of Hi-C data and compartment tracks reveals the emergence of series of shorter compartmental intervals in Δ*Nipbl* cells, with small B-like regions appearing inside A regions (Fig. 3a, b, Extended Data Fig. 7a). This finer compartmentalization is reflected in the shorter autocorrelation length of the compartment track: 150 kb in Δ*Nipbl* cells versus around 500 kb in wild-type and TAM cells (Extended Data Fig. 7b). The finer compartmentalization explains most of the remaining or new domains and boundaries seen in Δ*Nipbl* Hi-C maps (Extended Data Fig. 8a). These emerging B-like regions possess the hallmarks of compartmentalization: they are visible as local depressions in the Δ*Nipbl* compartment track (Fig. 3b) and they show preferential interactions with other B-regions in both far *cis*- and *trans*-chromosomal maps (Fig. 3b). As their diagonal squares lack both corner peaks and enriched borders, and exhibit mutual checkering, we conclude that these intervals do not represent newly formed TADs. By contrast, predominantly B-type regions of the genome do not show similar fragmentation in Δ*Nipbl* cells, despite a complete loss of TAD patterns (Extended Data Fig. 7a, b). Importantly, the appearance of finer compartmentalization upon *Nipbl* deletion shows that wild-type TADs can span regions of intrinsically opposing compartment type (Extended Data Fig. 8b–d).

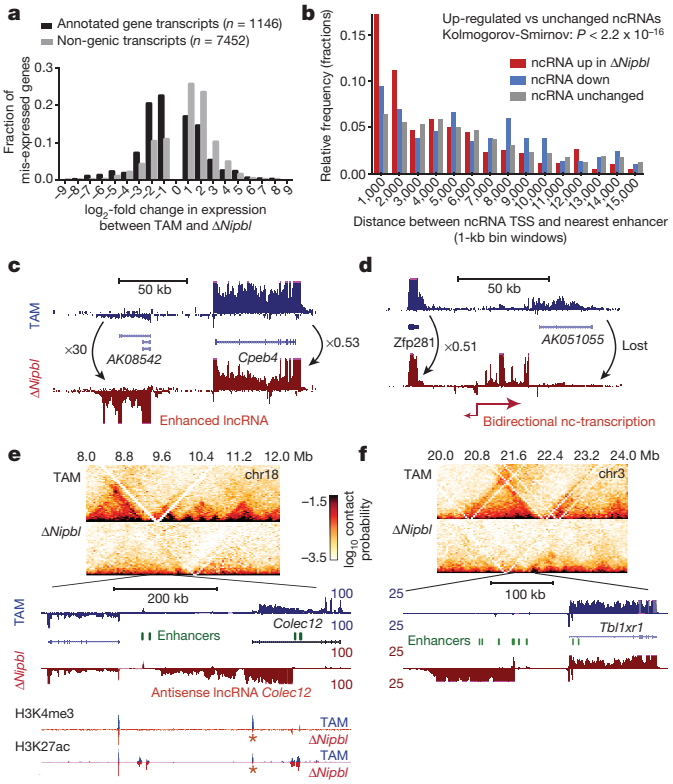


Figure 4 | Transcriptional changes in *Nipbl* mutants reveal possible enhancer–promoter miscommunication in absence of TADs. **a**, Distribution of fold-changes for genes (black, $n = 1{,}146$) and exogenic transcripts (light grey, $n = 7{,}452$) showing at least a twofold change in expression. **b**, **c**, Examples of transcriptional changes, with TAM control stranded RNA-seq tracks in blue, and Δ*Nipbl* in red. Four replicates per condition, each confirming the reported changes, were combined. lncRNA, long non-coding RNA; ncRNA, non-coding RNA. **d**, Distribution of distances between the transcriptional start of the ncRNAs (up $\log_2[\text{foldchange}] > 3$, $n = 595$; down $\log_2[\text{foldchange}] < -3$, $n = 284$; unchanged $-0.5 < \log_2[\text{foldchange}] < 0.5$, $n = 2{,}238$) and the nearest enhancer. **e**, **f**, Switches in transcription at the *Colec12* and *Tbl1xr1* loci. Top panels show 40-kb Hi-C maps of the 4-Mb neighbourhood. **e**, RNA-seq shows loss of *Colec12* transcripts, replaced by antisense transcripts initiated from an intronic enhancer. H3K4me3 and H3K27ac profiles show no changes at distal enhancers (green ovals), whereas peaks at the *Colec12* promoter disappear (red asterisk). **f**, Exogenic transcription emerges in Δ*Nipbl* cells upstream of the *Tbl1xr1* gene on chromosome 3.

Together, our observations defy the common notion of TADs simply being the building blocks of larger compartmental segments; instead, we conclude that TADs and compartments represent two independent, potentially antagonistic types of chromosomal organization.

Notably, we found that the compartmentalization profile of the Δ*Nipbl* Hi-C map reflects local transcriptional activity and chromatin state better than that of the wild-type Hi-C map (Fig. 3c, Extended Data Figs 7c, d, 8c, d). The compartment track of Δ*Nipbl* cells shows a stronger correlation with tracks of activity-associated epigenetic marks (for example, acetylation of H3K27 (H3K27ac), trimethylation of H3K4 (H3K4me3), DNase hypersensitivity, transcription factor binding), smoothed over a wide range of window sizes (Extended Data Fig. 7c). To understand the relationship between epigenetic state and the change in compartment status, we compared the compartment tracks to the mouse liver chromatin state segmentation (ChromHMM[31]) simplified into three state categories: active, repressed and inert (see Methods). Whereas inert regions are relatively unaffected by *Nipbl* deletion, regions of repressed and active chromatin further diverge in their compartment status (Fig. 3c, d), producing local peaks in the compartment track in active regions, and local B-like depressions in repressed regions (Fig. 3c). Furthermore, regions of facultative lamin-B1 association[32] are enriched in regions that show a reduction in compartment signal (from A-like to B-like), while those showing lamin-B1 association across different mouse cell lines are primarily B-type in both wild-type and mutant cells (Extended Data Fig. 9a, Methods). These changes in compartmentalization cannot be attributed to changes in expression or activity marks (H3K27ac and H3K4me3), which are largely unperturbed in Δ*Nipbl* cells at the scales relevant to compartmentalization (Extended Data Figs 8d, 9b–e). In summary, absence of cohesin enhances the compartmentalization of active and inactive chromatin as follows: (i) A and B regions, as detected by Hi-C, form fewer contacts between one another; (ii) a finer compartment division emerges; (iii) and this finer compartment structure corresponds better to the local functional states of the genome, even when considering those observed in wild-type cells. The fact that we observe no effect on activity marks indicates that cohesin and TADs do not play a role in the maintenance of epigenetic state, though this does not rule out possible roles in its establishment. A recent study has observed a similar preservation of epigenetic state upon CTCF depletion[18].

These results indicate that chromatin has an intrinsic tendency to segregate into compartments based on the local epigenetic landscape and transcriptional activity, and that Nipbl and cohesin activity interfere with this clear subdivision by bringing together and mixing loci with opposing states.

Transcriptional changes upon TAD loss

We next examined the effect of *Nipbl* deletion and disappearance of TADs on transcription. About a thousand genes show significant changes in expression (637 downregulated, median fold-change = 0.32; 487 upregulated, median fold-change = 3.15, with DESeq2 tools) in Δ*Nipbl* cells (Fig. 4a and Extended Data Fig. 10a–e). Gene ontology enrichment analysis does not give a strong indication of preferential effect on a biological function (Supplementary Table 4), reflecting possibly indirect and adaptive transcriptional changes.

While peaks of H3K27ac (and H3K4me3) at the promoters of affected genes change in coherence with expression changes, distal peaks (marking active distant enhancers) are mostly unaffected (Extended Data Fig. 10f–i), indicating that although transcriptional changes did occur, the regulatory potential of the cells was mostly unperturbed. There is so far no reliable way to identify *a priori* genes for which distal regulatory interactions are essential. However, we noticed that downregulated genes were surrounded by a larger intergenic space (defined by the distance separating the transcription start sites (TSSs) of their immediate neighbours) than upregulated or unaffected ones (Extended Data Fig. 10b) and that transcriptional changes were concentrated within regions that formed larger TADs in wild-type cells. This characteristic genomic context of transcriptional alterations is consistent with defective long-range regulatory interactions in Δ*Nipbl* cells.

Closer examination of RNA sequencing (RNA-seq) tracks revealed widespread upregulation of exogenic (intergenic or antisense intragenic) transcription in Δ*Nipbl* cells (Fig. 4a). While genes are more often downregulated than up-regulated in Δ*Nipbl* hepatocytes, we observed a clear opposite trend in exogenic regions, where transcription is more frequently upregulated than reduced. Using a conservative approach (see Methods), we found 1,107 non-genic transcripts or transcribed regions that showed at least an eightfold increase in transcription in Δ*Nipbl* cells; among these, 232 corresponded to non-coding RNAs, which were not or barely detected in control samples, and often not annotated (Extended Data Fig. 10d). The new transcription is often bi-directional, (Fig. 4c, d, Extended Data Fig. 10f–j) and occurs either at small pre-exiting H3K4me3 peaks, which are likely to correspond to poised promoters, or at active H3K27ac enhancers (Fig. 4b, e, f, Extended Data Fig. 10f–j). We saw several examples of reciprocal expression changes (that is, downregulation of a gene being followed by upregulation of an adjacent gene or of a new non-coding transcript) (Fig. 4c, d, Extended Data Fig. 10f–j), but often, new

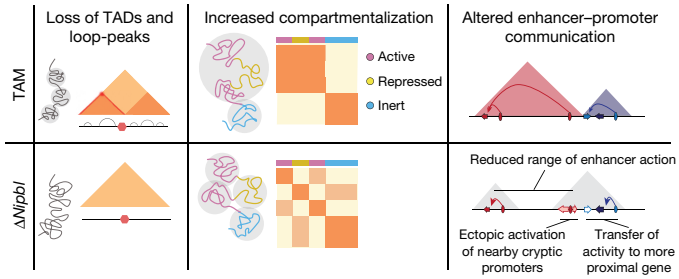


Figure 5 | Two independent but overlapping modes of chromosomal 3D organization. TADs (coloured triangles) and Hi-C peaks disappear upon *Nipbl* deletion (left), unmasking a stronger and finer compartmentalization (middle) that is visible as a fragmented checkered pattern in the Δ*Nipbl* Hi-C map relative to that of the wild type, and whose alternating member regions more faithfully track transcriptional activity. The resulting reduction in contact range (right) thwarts distant enhancers (ovals) from acting on their normal target genes (arrows, with coloured ones indicating active genes, white ones inactive), leading them to act instead on neighbouring genes or cryptic promoters located in their vicinity. The active units make up new compartmental regions (grey triangle).

non-coding transcription arose without measurable impact on surrounding genes. While the chromatin profile suggests that enhancers retain their normal activity and therefore regulatory potential, this rise in intergenic transcription initiated in the vicinity of distal regulatory elements suggests that *Nipbl* deletion impairs enhancer communication: with a reduced range of contact owing to the absence of TADs, some enhancers may not reach their target promoters and instead transfer their activity to nearby alternative, sometimes cryptic, targets (including themselves).

Cohesin is central to chromosome folding

Overall, our findings provide insights into the mechanisms that generate the 3D organization of the interphase genome and their relation to gene expression. Our data show that cohesin is essential for the formation of TADs. It is possible that in previous studies that did not report such drastic effects[19–21] cohesin depletion was insufficient to achieve substantial loss of TADs. Our simulations suggest that TADs would still be pronounced at twofold cohesin depletion, and that approximately eightfold depletion is required to induce loss of TADs; this is close to what we observed in Δ*Nipbl* samples. It is, however, also possible that *Nipbl* may affect cohesin activity at several levels, that is, not only as a loader but also by facilitating ATP hydrolysis and loop extrusion[33], which could further affect TAD formation. This could also account for why a recent deletion of *Mau2* (also known as *Scc4*), a co-factor of NIPBL, showed a milder effect on TAD formation[34].

While our manuscript was under review, several preprints reported direct perturbations to cohesin, producing results consistent with this study and thus providing additional support that loss of cohesin is responsible for the effects we observe. Single-nucleus Hi-C for cohesin (*Rad21*, also known as *Scc1*) knockout zygotes demonstrated complete loss of TADs and associated peaks[35]. However, the emergence of fine compartmentalization was not reported, probably owing to the limited number of nuclei studied. Two other preprints reporting degron-mediated depletion of *RAD21* in human cell lines observed concordant effects on TADs, peaks, and compartmentalization to our study[36,37]. In addition, one of these studies reported that cohesin depletion resulted in the emergence of 'loops' enriched in super-enhancers[37]. However, unlike the sharp, *cis*-only focal Hi-C peaks previously referred to as loops[12], these patterns manifest as grids of larger patches of contact enrichment appearing both in *cis* and in *trans*, and hence cannot be loops. Indeed, we also observed such features in our data using a new Hi-C browser[38], and we reported earlier that these patches simply appear to be strong compartmental interactions between very active regions and tend to line up well with transcribed genes[38].

Two independent folding principles

Our results challenge the classic picture of genome architecture in which TADs, peaks and compartments represent well-defined hierarchically folded entities, with individual TADs combining to form compartmental regions. Our observations show that there are at least two mechanisms of independent origin whose overlapping action organizes mammalian interphase chromatin (Fig. 5). The first mechanism, global spatial segregation and compartmentalization of the genome into active and inactive compartments, is achieved by a cohesin-independent mechanism, which acts globally and at varying scales, including scales smaller than previously appreciated. This compartmentalization is, however, suppressed by the action of a second, cohesin-dependent, mechanism that compacts chromatin locally, independently of its status, resulting in formation of TAD patterns in Hi-C maps. Notably, the independence of the two mechanisms is supported by the absence of compartments, despite the existence of TADs, in maternal zygotic pronuclei[39]. Local compaction can be achieved by a cohesin-driven and energy-dependent loop extrusion mechanism, akin to the motor function recently observed in another SMC complex, condensin[40]. Another recent study[41] demonstrated that active loop extrusion in a polymer model of chromatin can indeed lead to reduced compartmentalization and failure of short compartmental segments to segregate, as observed here.

The co-existence of two processes with different modes and scales of action can help to explain the difficulties in the field in delineating and unambiguously classifying the different features of Hi-C maps, leading to a confusing plethora of denominations and definitions[7,8,12,42–44]. Our data clearly demonstrate the existence of local, cohesin-dependent, self-interacting domains identifiable as TADs. The experimental ability to distinguish between the two modes of chromosome organization should enable investigations of the process(es) that govern their formation and maintenance, as well as characterization of their relationships to gene expression. In this respect, while compartmental segregation may facilitate the maintenance of regulatory interactions, the ability of cohesin to counteract segregation and bring regions of different activities together may play an essential role in initiating changes in gene expression driven by distant enhancers.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 December 2016; accepted 19 September 2017.
Published online 27 September 2017.

1. Bickmore, W. A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152,** 1270–1284 (2013).
2. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62,** 668–680 (2016).
3. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164,** 1110–1121 (2016).
4. Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat. Rev. Genet.* **14,** 390–403 (2013).
5. Nora, E. P., Dekker, J. & Heard, E. Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods? *BioEssays* **35,** 818–828 (2013).
6. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326,** 289–293 (2009).
7. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485,** 376–380 (2012).
8. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485,** 381–385 (2012).
9. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161,** 1012–1025 (2015).
10. Dowen, J. M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159,** 374–387 (2014).
11. Symmons, O. *et al.* The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. *Dev. Cell* **39,** 529–543 (2016).
12. 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).

13. Merkenschlager, M. & Nora, E. P. CTCF and cohesin in genome folding and transcriptional gene regulation. *Annu. Rev. Genomics Hum. Genet.* **17,** 17–43 (2016).
14. Fudenberg, G. *et al.* Formation of chromosomal domains by loop extrusion. *Cell Reports* **15,** 2038–2049 (2016).
15. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112,** E6456–E6465 (2015).
16. Guo, Y. *et al.* CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* **162,** 900–910 (2015).
17. de Wit, E. *et al.* CTCF binding polarity determines chromatin looping. *Mol. Cell* **60,** 676–684 (2015).
18. Nora, E. P. *et al.* Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169,** 930–944 (2017).
19. Zuin, J. *et al.* Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl Acad. Sci. USA* **111,** 996–1001 (2014).
20. Seitan, V. C. *et al.* Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments. *Genome Res.* **23,** 2066–2077 (2013).
21. Sofueva, S. *et al.* Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32,** 3119–3129 (2013).
22. Nasmyth, K. & Haering, C. H. Cohesin: its roles and mechanisms. *Annu. Rev. Genet.* **43,** 525–558 (2009).
23. Tedeschi, A. *et al.* Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501,** 564–568 (2013).
24. Gerlich, D., Koch, B., Dupeux, F., Peters, J.-M. & Ellenberg, J. Live-cell imaging reveals a stable cohesin-chromatin interaction after but not before DNA replication. *Curr. Biol.* **16,** 1571–1578 (2006).
25. Hu, B. *et al.* Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res.* **43,** e132 (2015).
26. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30,** 90–98 (2011).
27. Schmitt, A. D. *et al.* A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Reports* **17,** 2042–2059 (2016).
28. Lajoie, B. R., Dekker, J. & Kaplan, N. The Hitchhiker's guide to Hi-C analysis: practical guidelines. *Methods* **72,** 65–75 (2015).
29. Naumova, N. *et al.* Organization of the mitotic chromosome. *Science* **342,** 948–953 (2013).
30. Nolen, L. D., Boyle, S., Ansari, M., Pritchard, E. & Bickmore, W. A. Regional chromatin decompaction in Cornelia de Lange syndrome associated with NIPBL disruption can be uncoupled from cohesin and CTCF. *Hum. Mol. Genet.* **22,** 4180–4193 (2013).
31. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).
32. Meuleman, W. *et al.* Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **23,** 270–280 (2013).
33. Rhodes, J., Mazza, D., Nasmyth, K. & Uphoff, S. Scc2/Nipbl hops between chromosomal cohesin rings after loading. *eLife* **6,** e30000 (2017).
34. Haarhuis, J. H. I. *et al.* The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169,** 693–707 (2017).
35. Gassler, J. *et al.* A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. Preprint at https://www.biorxiv.org/content/early/2017/08/17/177766 (2017)
36. Wutz, G. *et al.* CTCF, WAPL and PDS5 proteins control the formation of TADs and loops by cohesin. Preprint at https://www.biorxiv.org/content/early/2017/08/18/177444 (2017).
37. Rao, S. *et al.* Cohesin loss eliminates all loop domains, leading to links among superenhancers and downregulation of nearby genes. Preprint at https://www.biorxiv.org/content/early/2017/05/18/139782 (2017).
38. Kerpedjiev, P. *et al.* HiGlass: web-based visual comparison and exploration of genome interaction maps. Preprint at https://www.biorxiv.org/content/early/2017/03/31/121889 (2017).
39. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-to-zygote transition. *Nature* **544,** 110–114 (2017).
40. Terakawa, T. *et al.* The condensin complex is a mechanochemical motor that translocates along DNA. *Science* http://dx.doi.org/10.1126/science.aan6516 (2017).
41. Nuebler, J. *et al.* Chromatin organization by an interplay of loop extrusion and compartmental segregation. Preprint at https://www.biorxiv.org/content/early/2017/10/03/196261 (2017).
42. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153,** 1281–1295 (2013).
43. Fraser, J. *et al.* Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol. Syst. Biol.* **11,** 852 (2015).
44. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148,** 458–472 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work would not have been possible without the contribution of the members of the EMBL Laboratory Animal Resources Facility, particularly S. Feller, to animal welfare and husbandry. We thank members of the Mirny and the Spitz labs, J. Marioni (EMBL/EBI)) for discussions and suggestions, H. Marlow for help and expertise in generating the cohesin and CTCF ChIP–seq data, the EMBL Genomics Core Facility and Pasteur Biomics for assistance and support in sequencing the genomic libraries, and A. Losada for providing antibodies. W.S. and A.P. were supported by an EMBL Interdisciplinary Postdoc (EIPOD) Fellowship under Marie Curie Actions COFUND. The work in the Mirny laboratory is supported by R01 GM114190, U54 DK107980 from the NIH, and 1504942 from the NSF. The collaboration is also partially supported by the MIT-France MISTI Fund. The work in the Spitz laboratory was supported by EMBL, the Pasteur Institute and the Deutsche Forschungsgesellschaft (DFG grant: SP 1331/3-1) and included funding from the European Commission's Seventh Framework Programme through the Collaborative Research Project RADIANT (grant agreement no. 305626, to W.H.).

Author Contributions W.S. and F.S. conceived the study and designed the experiments, with input and advice from C.H.H. Experimental data were generated by W.S., with the help of A.P. for TCC; A.P. also carried out preliminary analyses of the TCC datasets, with advice from W.H. N.A. and A.G. performed computational analysis of Hi-C, RNA-seq, ChIP–seq and other relevant datasets. Y.L.-M. and N.A.F. contributed to analysis of transcription data. G.F. performed computer simulations of cohesin depletion and assisted with data analysis and paper writing. L.M. provided advice on data analysis and simulations. W.S., N.A., A.G., L.M. and F.S. wrote the paper with input from the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to F.S. (francois.spitz@pasteur.fr) or L.M. (leonid@mit.edu).

Reviewer Information *Nature* thanks K. Shirahige and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

Experimental procedures. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Generation of *Nipbl*$^{flox/flox}$ mice. The *Nipbl* locus was targeted by introduction of two *loxP* sites flanking exon 18 via homologous recombination in E14 mouse embryonic stem (ES) cells. Individual ES cell clones were screened for successful recombination by Southern blotting with two unique probes that hybridized 5′ and 3′ to the integration site. Cells of a successful clone were injected into mouse blastocysts and the resultant chimaera were bred to C57BL/6J mice. The offspring were analysed for successful germ line transmission by PCR and Southern blotting. *Nipbl*$^{flox/+}$ mice were maintained on a C57BL/6J genetic background.

For acute deletion of the *floxed* exon, we used either a constitutive ubiquitous Cre-driver (Hprt::Cre)[45], a limb-specific Cre-driver (Prx1::Cre)[46] or an inducible, liver-specific *cre* allele (*Ttr-cre*/Esr1)[47].

Mice were genotyped by PCR using specifc primer pairs (details available on request). All lines were maintained by breeding with C57Bl/6J mice. Mouse experiments were conducted in accordance with the principles and guidelines in place at the European Molecular Biology Laboratory, as defined and overseen by its Institutional Animal Care and Use Committee, in accordance with the European Directive 2010/63/EU. The experimental protocols followed were reviewed and approved by the EMBL Institutional Animal Care and Use Committee under the project 'Function of cohesin and condensing complexes'.

Generation and preparation of *Nipbl*$^{-/-}$ adult primary hepatocytes. To inactivate *Nipbl* in adult mouse hepatocytes, *Nipbl*$^{flox/+}$ mice were crossed with mice carrying an inducible, liver-specific *cre* allele (*Ttr-cre*/Esr1)[47]. The resultant double heterozygous mice were backcrossed to *Nipbl*$^{flox/flox}$ mice. For experiments we used animals homozygous for the floxed *Nipbl* allele carrying either one or no copy of the inducible, liver-specific *cre* allele as sample (*Ttr-cre/Esr1*$^{+/wt}$; *Nipbl*$^{flox/flox}$) or control (*Ttr-cre/Esr1*$^{wt/wt}$; *Nipbl*$^{flox/flox}$) mice, respectively.

Twelve-week-old mice were injected with 1 mg tamoxifen (100 µl of 10 mg/ml tamoxifen in corn oil) on five consecutive days. After another five days without injection, the mice were killed and the hepatocytes harvested. Until this time point, mice displayed no abnormal behaviour, weight loss or any other obvious physiological changes. This was also the case when we kept mice for additional four days without injection to test for any adverse effects immediately after our experimental time point.

The liver was dissected and the left lateral lobe was prepared for a two-step perfusion by an adapted method[48,49]. First, the liver was perfused with an EDTA-containing buffer to remove Ca^{2+} from the tissue. This weakens the integrity of the desmosome, which is then digested during the subsequent perfusion with a Ca^{2+} rich buffer containing collagenase. The freed hepatocytes were rinsed through a cell strainer and washed four times with ice-cold Ca^{2+}-rich buffer without collagenase. For each wash the cells were spun at low centrifugal force (60*g* for 1 min) to reduce non-mesenchymal cells and debris and thereby enrich intact hepatocytes in the sample. Part of each sample was fixed with 1% PFA for 10 min at room temperature. Fixed and unfixed hepatocytes were aliquoted and frozen in liquid N_2 for later use.

***Nipbl* RNA levels and activity.** Unfixed hepatocyte aliquots were thawed and RNA was prepared with Qiagen RNeasy Kit. cDNA was generated using NEB ProtoScript First Strand cDNA Synthesis Kit with random primer mix. RT–qPCR was performed with Applied SYBR Green PCR Master Mix and the following primers: *Nipbl*-qPCR_F TCCCCAGTATGACCCTGTTT, *Nipbl*-qPCR_R AGAACATTTAGCCCGTTTGG, *Gapdh*-qPCR_F CTCCCACTCTTC CACCTTCG, *Gapdh*-qPCR_R CCACCACCCTGTTGCTGTAG, RTqPCR_*Pgk1*_ Fwd TGGTATACCTGCTGGCTGGATGG and RTqPCR_*Pgk1*_Rev GACC CACAGCCTCGGCATATTTC.

For western blots, unfixed hepatocyte aliquots were lysed and fractionated with a Subcellular Protein Fractionation Kit (ThermoFisher). The blots were probed with antibodies against cohesin subunits SA-1 and SMC1 (from A. Losada, CNIO, Madrid) and Topo IIβ (611492, BD Biosciences) and histone H2B (07-371, Millipore) as controls for nuclear soluble and nuclear insoluble fractions, respectively.

Immunohistochemistry on liver. Slices of adult liver were collected and fixed in 4% PFA overnight. After dehydration, the tissues were embedded in paraffin and sectioned at 6 µm. The sections were deparaffinized with xylene and rehydrated, and antigens were retrieved by boiling in citrate buffer. The sections were blocked in 10% FBS and incubated with primary antibodies (anti-phospho-histone H3, 06-570 Millipore; anti-cleaved-caspase-3, #9661 Cell Signaling) at 4 °C overnight. Primary antibodies were detected with goat anti-rabbit IgG Alexa Fluor 568 secondary antibody (A-11011, Invitrogen) and counter stained with DAPI. Images were acquired using confocal microscopy.

RNA-seq libraries and sequencing. RNA integrity was tested with Bioanalyzer (Agilent RNA Nano Kit) and ribosomal RNA was removed using Ribo-Zero rRNA Removal Kit (Illumina) before library preparation. Strand-specific libraries were prepared with NEBNext Ultra Directional RNA Library Prep Kit for Illumina. After amplification and size selection with Agencourt AMPure XP beads (Beckmann Coulter), size distributions were determined with Bioanalyzer. Equimolar pools of libraries were sequenced with Illumina HiSeq2000 (50 bp, single end). We retrieved on average 25 million reads per sample, of which 19 million reads were uniquely mapped to the reference genome (NCBI37/mm9).

ChIP–seq libraries and sequencing. For histone ChIP–seq, fixed aliquots of hepatocytes were hypotonically lysed and sonicated in 1% SDS/TE. An aliquot of each sample was reverse cross-linked in order to determine chromatin concentration and sonication efficiency. Twenty micrograms of chromatin per sample was diluted in RIPA and incubated with 1.5 µg of either anti-H3k4me3 antibody (C15410003-50, Diagenode) or anti-H3K27Ac antibody (ab4729, Abcam) at 4 °C overnight. The antibodies were retrieved with Dynabeads (IgA, Invitrogen) and bound chromatin was washed and eluted. After reverse cross-linking, the amount of immunoprecipitated and input DNA was determined with Qubit (Thermo Fisher). The libraries were prepared with NEBNext ChIP–Seq Library Prep Kit for Illumina. After amplification and size selection with E-Gel SizeSelect (Thermo Fisher) the size distributions were determined with Bioanalyzer. Equimolar pools of libraries were sequenced with Illumina HiSeq2000 (50 bp, single end). We retrieved on average 20 million reads per sample, of which 15 million reads were uniquely mapped to the reference genome (NCBI37/mm9).

For cohesin subunits and CTCF ChIP–seq, cell lysis, isolation of nuclei and shearing of chromatin were performed as described[50]. In brief, cross-linked samples were thawed on ice and incubated in lysis buffer (5mM PIPES pH 8.0, 85mM KCl, 0.5% IGEPAL CA-630) supplemented with complete, EDTA-free protease inhibitor cocktail (Sigma) for 45 min before being transferred to a shearing cuvette (Covaris-520130). Lysis was performed in a Covaris E220 sonicator for 5 min according to the following parameters: peak power = 75 W, duty factor = 2%, 200 cycles per burst, temperature = 4 °C. Following lysis, nuclei were pelleted in a chilled 4 °C centrifuge at 1,000*g* for 5 min and resuspended in shearing buffer (10mM TrisHCl pH 8.0, 0.1% SDS, 1mM) EDTA supplemented with complete, EDTA-free protease inhibitor cocktail (Sigma) and transferred to a clean shearing cuvette. Chromatin was sheared in a Covaris E220 sonicator for 17 min with the following settings: peak power = 140 W, duty factor = 5%, 200 cycles per burst, temperature = 4 °C. An aliquot of 1% of the total chromatin lysate was reverse cross-linked and subsequently used to validate chromatin fragment size distribution, concentration and for the construction of an input control library. The antibodies used (anti-Rad21: Abcam ab992; anti-CTCF: Millipore 07-729; anti-SMC3: Abcam ab9263) were raised against epitotes fully conserved between the mouse and human proteins. Each antibody was incubated overnight with 100 µl of Dynabeads M-280 sheep anti-rabbit. ChIP was performed successively (Rad21, followed by CTCF, followed by SMC3) using unbound chromatin from the first immunoprecipitation to perform the next one. As immunoprecipitation efficiency is very low for this protein, we did not observe a change of signal when comparing sequential to parallel immunoprecipitation performed on control chromatin. Antibody-bound beads were washed twice in ice-cold PBS and incubated with 75 µg sheared mouse hepatocyte chromatin pooled with 75 µg HEK293 human chromatin (as internal control and calibration) in RIPA buffer (50 mM TrisHCl, pH 8.0, 0.15 M NaCl, 1% Triton X-100 and 0.1% sodium deoxycholate) overnight at 4 °C on a rotating wheel and then subsequently washed twice each in RIPA, RIPA-500 (50mM TrisHCl, pH 8.0, 0.5M NaCl, 1% Triton X-100 and 0.1% sodium deoxycholate), LiCl buffer (50mM TrisHCl, pH 8.0, 1mM EDTA, 1% IGEPAL CA-630, 0.7% sodium deoxcholate and 0.5% LiCl), and TE for 5 min per wash on a rotating wheel. Beads were resuspended in TE with 50 µg/ml RNaseA and incubated for 30 min at 37 °C, followed by reverse cross-linking with 0.5 mg/ml proteinase K in TE supplemented with 0.5% SDS overnight at 65 °C in a shaking thermomixer at 1,200 r.p.m. DNA was purified with the QIAquick nucleotide removal kit (Qiagen 28304). Multiplexed libraries were prepared according to the NEBNext ChIP–seq library preparation protocol and amplified with 15 PCR cycles using Thermo Phusion HF Mastermix (Fisher 10402678). Sequencing was performed on a NextSeq500 sequencer using a HighOutput Kit v2 75 (Illumina FC-404-2005).

HEK293 cells, chromatin from which is used as internal control for the immunoprecipitation, were not authenticated, nor screened for mycoplasma contamination.

Tethered chromatin capture. Roughly 100 million fixed hepatocytes per sample were processed as described[26] using HindIII. Libraries were PCR-amplified (12 cycles) and size selected with E-Gel SizeSelect (Thermo Fisher). Equimolar pools of libraries were sequenced with Illumina HiSeq2000 (50 bp, paired end). We retrieved between 100 and 150 million paired reads per sample, of which ~40% had both sides uniquely mapped to the reference genome (NCBI37/mm9).

Computational analysis: preparation of Hi-C maps. We mapped the sequence of Hi-C molecules to reference mouse genome assembly mm9 using Bowtie 2.2.8 and the iterative mapping strategy as described[51], implemented in the *hiclib* library

for Python (https://bitbucket.org/mirnylab/hiclib). Upon filtering PCR duplicates and reads mapped to multiple or zero locations, we aggregated the reads pairs into 20-kb and 100-kb genomic bins to produce Hi-C contact matrices.

Filtering of Hi-C maps. Low-coverage bins were excluded from further analysis using the MAD-max (maximum allowed median absolute deviation) filter on genomic coverage. MAD-max is a robust heuristic filtering method based on the empirical observation that the total number of contacts per genomic bin follows a log-normal distribution. This shape of a distribution is expected when different factors (GC content, mappability, restriction fragment density and so on) affect the visibility of a genomic bin in a multiplicative fashion. The MAD-max procedure has three steps:

1. Normalize the number of contacts at each genomic bin by dividing them by their corresponding chromosome-wide medians. This step accounts for possible variation in chromosome copy numbers, for example, in the X chromosome.

2. Fit the distribution of normalized contact numbers with a log normal distribution. To make this step more robust against skew caused by heavy tails, we find the centre of the distribution as the median of log[normalized contact numbers], and the width as the median absolute deviation (MAD) from this median.

3. Filter out the low-coverage genomic bins, which are 3 s.d. below the centre of the distribution. The exact conversion ratio between MAD and s.d. is 1.4826 and we approximate 3.0 standard deviations with 5.0 MADs.

Our MAD-max filtering procedure removed between 1.4% and 2.2% of non-zero bins in the six experimental replicates processed at 20-kb resolution and between 1.8% to 3.0% at 100-kb resolution.

To remove the short-range Hi-C artefacts—unligated and self-ligated Hi-C molecules—we removed contacts mapping to the same or adjacent genomic bins. The filtered 20-kb and 100-kb contact matrices were then normalized using the iterative correction procedure[51], such that the genome-wide sum of contact probability for each row or column equals 1.0. Observed and expected contact maps were obtained by dividing each diagonal of a contact map by its average value over non-filtered genomic bins.

The same procedure was used to analyse other existing Hi-C data sets on cohesin-depleted cells[19–21].

Compartment analysis by eigenvector decomposition. The compartment structure of Hi-C maps was detected using a modified procedure[51]. In short, compartment tracks (that is, the values of compartment signal across all genomic bins) were quantified as the dominant eigenvector of the observed or expected 20-kb and 100-kb *cis* contacts maps upon subtraction of 1.0 and rescaling by the magnitude of the square root of its eigenvalue, as implemented in *hiclib*. We refer to continuous genomic regions of relatively uniform compartment signal as compartment intervals. We introduce the genome-wide measure of a Hi-C map compartmentalization as $\sqrt{\mathrm{AA} \times \mathrm{BB}/\mathrm{AB}^2}$, where AA is the average contact enrichment between pairs of loci with a strong compartment A signal (those with eigenvector values from the top 20% of the genome-wide range of eigenvector values), BB is the average contact enrichment between pairs of loci with a strong compartment B signal (those with eigenvector values from the bottom 20% of the genome-wide range of eigenvector values) and AB is the average contact depletion between pairs of loci with a strong compartment A and B signal.

Domain detection (Lavaburst) and peak coordinates. To identify domains, we used a segmentation algorithm very similar to one previously published[52], which divides the genome into domains in such a way as to maximize a global domain scoring function. We used two different scoring functions: one was the corner score function[12] and the other was based on network modularity[53], which is a metric widely used to detect communities in networks. The modularity score for a domain spanning genomic bins a to b inclusively is given by

$$S(a,b) = \sum_{i=a}^{b}\sum_{j=i}^{b} A_{ij} - \gamma N_{ij},$$

where A is the contact matrix and N is the corresponding matrix of a penalizing background model. The resolution parameter γ controls the strength of the penalty and therefore the characteristic size of the domains identified.

By restricting the solution space to contiguous segmentations, both calculating domain scores and finding the highest scoring segmentation can be reduced to $O(n^2)$ dynamic programming algorithms. Optimal segmentation, in particular, becomes the well-known max-sum algorithm on a weighted directed acyclic graph[54]. Furthermore, one can marginalize over the space of all possible segmentations to obtain a linear genomic track of boundary scores for a given domain scoring function, also in $O(n^2)$ time. The implementation of these and related algorithms is provided in the lavaburst package (https://github.com/nvictus/lavaburst).

The domain calls used in composite heatmaps were computed using the corner score on 20-kb resolution matrices. To robustly call insulating boundaries across different conditions, we exploit the multi-resolution nature of the modularity score and compute the average marginal boundary scores on 20-kb wild-type and Δ*Nipbl* contact matrices sweeping over a range of γ values to obtain a one-dimensional boundary (that is, insulation) track. Short intervals representing insulating loci were called by thresholding on the boundary score, and the common and unique loci to each condition were determined by interval intersection.

To characterize the structure of known peaks in our data, we used the list of peaks detected in Hi-C maps for the CH12-LX mouse cell line[12].

Analysis of epigenetic features. To study the influence of the epigenetic chromatin state on genomic architecture, we used the ChromHMM genomic state annotation[55]. In brief, the authors trained the ChromHMM genomic segmentation model on H3K4me1, H3K4me3, H3K36me3, H3K27me3, H3K27ac, CTCF and RNA polymerase II ENCODE tracks for mouse liver. This model assigned to each genomic locus one of 15 possible epigenetic states. For our analyses, we further grouped these 15 states into three: active (states 1–9, 14 and 15 characterized by presence of PolII), repressed (states 10–12, characterized by high H3K27me3 and low H3K27ac emission probabilities) and inert (state 13, lack of any signal).

To find the average footprint of a chromatin-bound CTCF, we detected all occurrences of the M1 CTCF motif[56] in the mm9 genome (filtered by $P > 0.0001$), intersected with CTCF binding peaks in ENCODE adult mouse liver data set[57]. This procedure yielded 27,840 peaks.

To study how lamina association affects genome compartmentalization, we used a data set of Lamin-B1-binding loci[58] containing data from four mouse cell lines: ES cells, multipotent neural precursor cells (NPCs), terminally differentiated astrocytes and embryonic fibroblasts (MEFs). We then selected 100-kb genomic bins with more than 60 locations probed for lamin-B1 binding. We called the bins as constitutive LAD (cLAD) or non-LAD bins if >90% of probed locations were bound to lamina or not bound to lamina across all four tested cell lines, respectively. Finally, bins were called facultative LADs if more than 90% of probes showed lamina binding in some cell lines, but not in others.

RNA-seq. We mapped the RNA-seq data to the mm9 reference mouse genome assembly and GENCODE vM1 transcriptome[59] using STAR v.2.5.0a[60] and scripts from the ENCODE RNA-seq pipeline (https://github.com/ENCODE-DCC/long-rna-seq-pipeline). To obtain the tracks of local transcription, we aggregated the uniquely mapped reads into RPM-normalized bigWig files using the built-in STAR functionality. To find differentially expressed genes, we aggregated the read counts at the gene level using HTSeq[61] with the 'union' option and called differentially expressed genes with DESeq2[62].

To identify exogenic transcripts, we aligned reads on the mm9 genome using HISAT2[63] with the option to authorize splicing events. From these alignments, we counted the number of reads in sliding windows of 1 kb (with a step of 600 nt) taking each strand of the genome as separate. We use a Gaussian mixture model (two distributions) to find a cutoff value for readcounts differentiating 'expressed' regions from noise. We choose a cutoff of 15 reads according to the best-fitted Gaussian distributions. To avoid overestimating the number of exogenic transcripts, we merged adjacent expressed windows (on the same strand) as composite transcripts. We used spliced reads to merge expressed windows that were not directly adjacent, but linked by spliced reads, with a cutoff of seven splice reads. The resulting merged and linked windows defined the region corresponding to the exogenic transcription unit, and we took its 5′ end (using the strand information, and after trimming the 5′ end of all nucleotides with no coverage) as the transcriptional start site.

ChIP–seq. We mapped the sequences obtained in ChIP–seq experiments using workflows based on the ENCODE 3 pipeline (https://github.com/ENCODE-DCC/chip-seq-pipeline).

For the CTCF and cohesin (Rad21 and SMC3) ChIP experiments, we aligned the reads from the pooled experimental and calibration (human spike-in) samples using *bwa mem* to a combined reference consisting of the mm9 and hg19 genome assemblies. For each condition, the reads that mapped uniquely to one of the two assemblies were bucketed into two separate data sets. The remainder of the ChIP–seq processing pipeline was applied to both the hg19-mapped and mm9-mapped reads in order to generate raw signal tracks. Because the proportion of human spike-in DNA was the same in both conditions, we used the difference in hg19 read coverage between the TAM control and Δ*Nipbl* conditions to 'calibrate' the raw mm9 signal pileups in order to make the binding profiles quantitatively comparable. In brief, for each factor (CTCF, Rad21, SMC3), we computed genome-wide binned coverage at 200 bp of the hg19-mapped reads. Because the background signal that makes up most of the genome-wide track exhibits different signal-to-noise characteristics than the binding peaks whose signal we wish to compare (as can be surmised from Extended Data Fig. 6b), we used the ratio of the top 0.2% of 200-bp bins in TAM vs Δ*Nipbl* to rescale the mm9 signal in the Δ*Nipbl* condition in order to be compared with the mm9 signal in the TAM condition. Peak calling was performed using MACS2 on each of the mm9 data sets using alignments of mm9-mapped input DNA as background.

GO enrichment analysis. Gene ontology enrichment analysis was performed on significantly differentially expressed genes (DESeq2 adjusted $P < 0.5$) using the online Functional Annotation Tool from DAVID version 6.8 (http://david.ncifcrf.gov), selecting annotation terms from the GOTERM_BP_DIRECT, GOTERM_MF_DIRECT, and GOTERM_CC_DIRECT categories and using default parameters. For the background gene list, we took the subset of gene IDs reported by DESeq2 that were expressed in either the TAM control condition or the Δ*Nipbl* condition. Extended Data Table 1 displays the list of enriched GO terms with a Benjamini–Hochberg $P < 0.05$.

Simulations of loop extrusion. To investigate the impact of NIPBL depletion on TADs and associated peaks in the context of loop extrusion, we performed simulations that couple the 1D dynamics of loop extrusion by LEFs with 3D polymer dynamics, as described previously[14]. We then generated contact maps and calculated $P(s)$, also as previously described[14].

We considered a system of 42 consecutive TADs of 400 kb, where impermeable boundary elements were placed between each pair of neighbouring TADs. We modelled the chromatin fibre as a series of 1-kb monomers (~6 nucleosomes, ~20 nm), such that each TAD was 400 monomers. Polymer connectivity, stiffness, and excluded volume were implemented as previously described[14]. Extruded loops held by LEFs were implemented by connecting the two monomers held by the two heads of each LEF with a harmonic bond, as previously described[14,64].

Three-dimensional polymer dynamics were implemented using OpenMM, a high-performance GPU-assisted molecular dynamics API[65,66]. Simulations were initialized from a compact polymer conformation as described[67], created on a cubic lattice a box of the size (PBC box – 2). Prior to a block of simulations, LEF dynamics were advanced by 500,000 steps. To allow the polymer fibre to equilibrate with this set of LEF-imposed bonds, simulations were then advanced for 400 blocks of simulations. After that, 2,000 blocks (that is, blocks of 1D+3D dynamics) of simulations were performed and their conformations were recorded. After that, LEF dynamics were advanced by 500,000 steps, and the process was repeated, until 5,000 conformations for each parameter set were obtained.

For the control simulation, we considered LEFs with 200-kb processivity and 400-kb separation. To investigate the effect of depleting the amount of bound cohesin, we then increased LEF separation by twofold, fourfold, and eightfold. All simulations used a stiffness of 2, density of 0.2, and 3D-to-1D-steps of 2,500. To generate contact maps from simulated conformations, a capture radius of 6 was used.

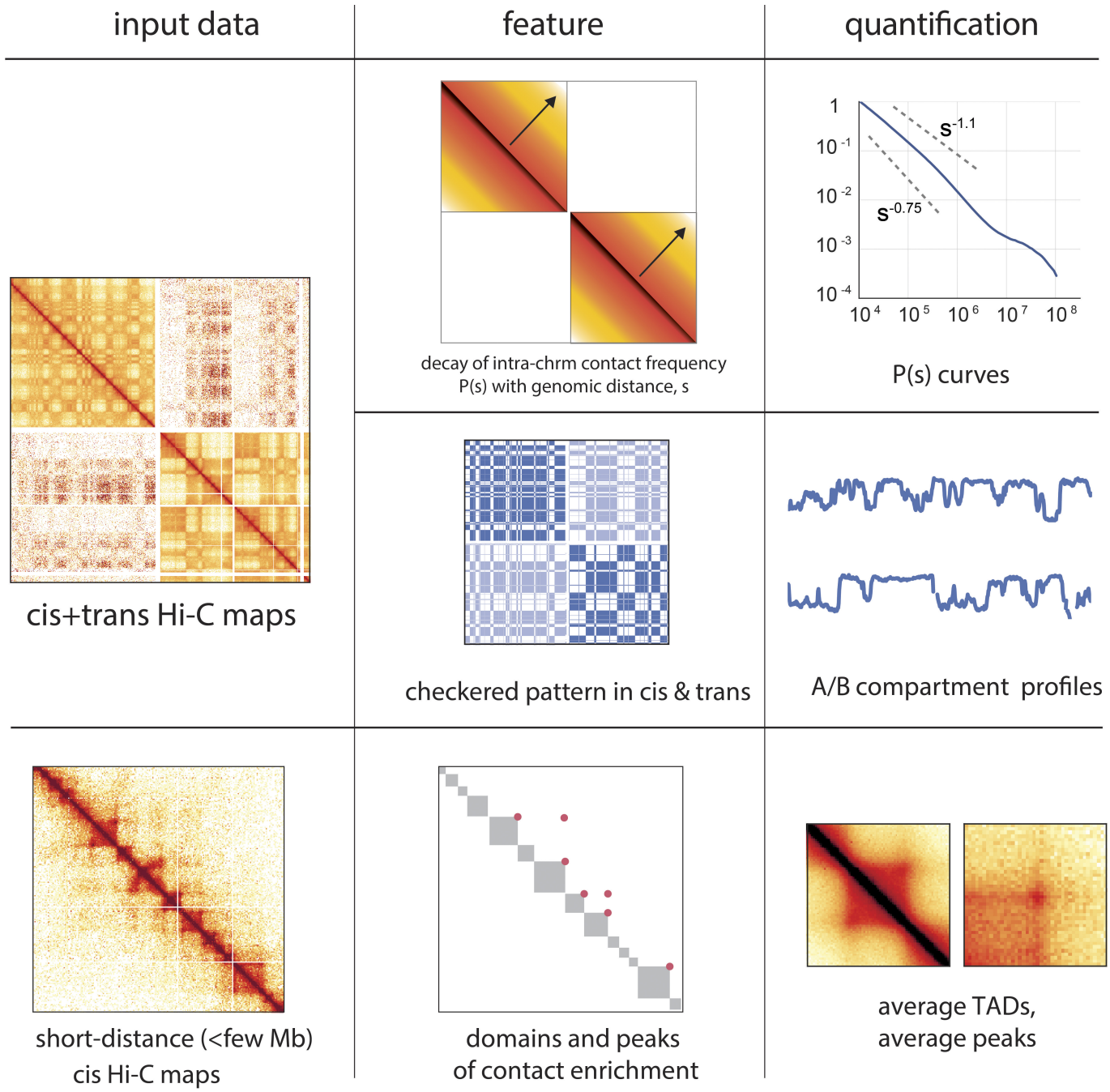
For display, simulated contact maps were first binned using 10 monomer (10-kb) bins. Then, the map was normalized such the average value of the first diagonal equals one, and $\log_{10}$ transformed. Finally, the colour-scale was clipped to show 1.5 logarithmic orders of magnitude as its dynamic range.

To calculate $P(s)$ plots from simulated data, a contact map with 1-monomer resolution was used. For each diagonal of the contact map, we determined which regions within 400-monomer TADs and which were between TADs. We then averaged the values in logarithmically-spaced bins of increasing distance with a step of 1.1. Similar to experimental data, $P(s)$ curves were then vertically shifted such that $P(s)$ at 10 kb was equal to 1.

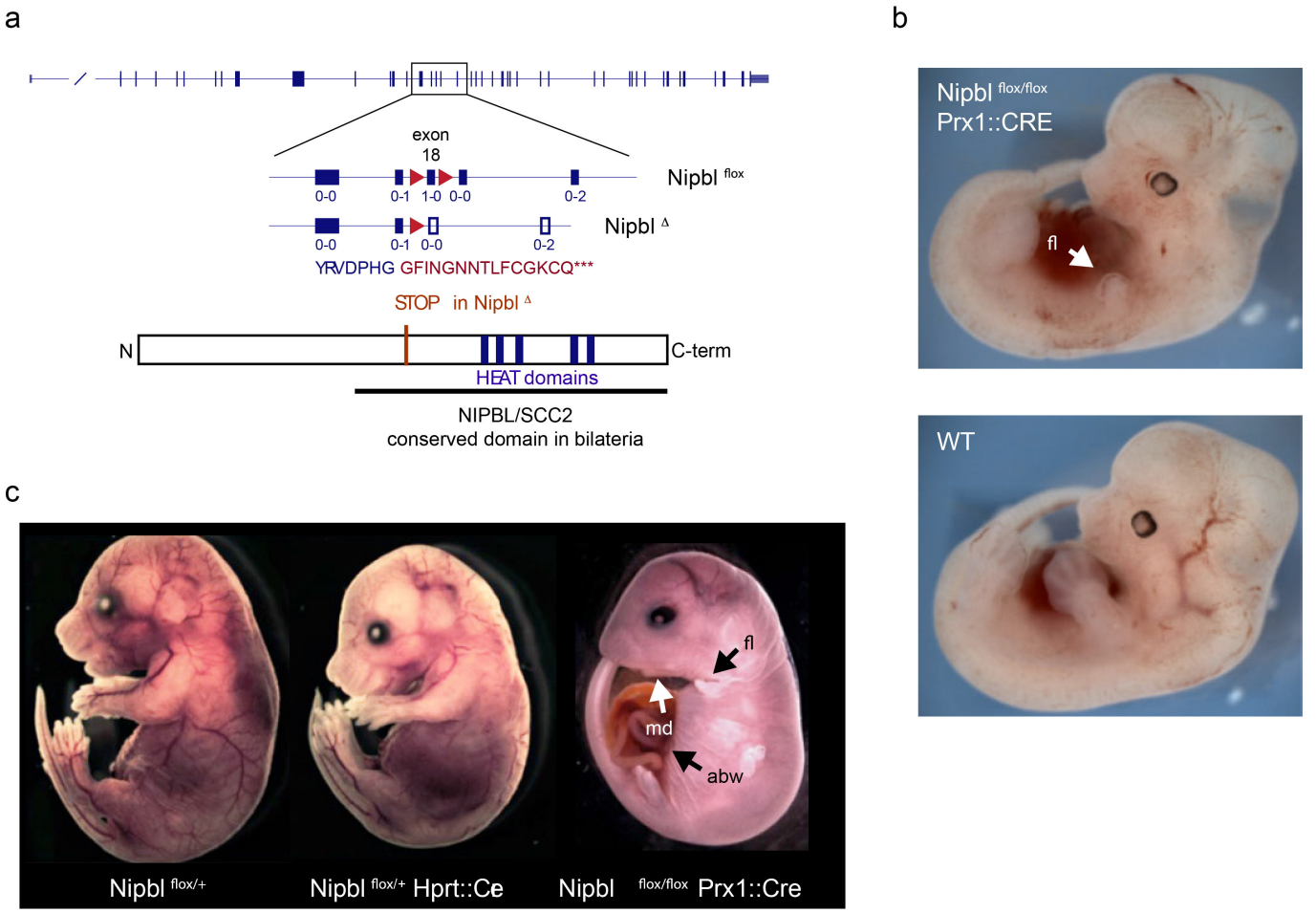
Code availability. The software developed to process the data in our study is available at https://github.com/mirnylab/.

Data availability. The reference and accession numbers of published data used and analysed in this work are indicated in Supplementary Table 3. The Hi-C, ChIP–seq and RNA-seq data generated and analysed are available in the GEO repository, under accession number GSE93431. Links to HiGlass visualizations of Hi-C, ChIP–seq and RNA-seq data obtained in this study are available at http://mirnylab.mit.edu/projects/nipbl and at http://higlass.io.

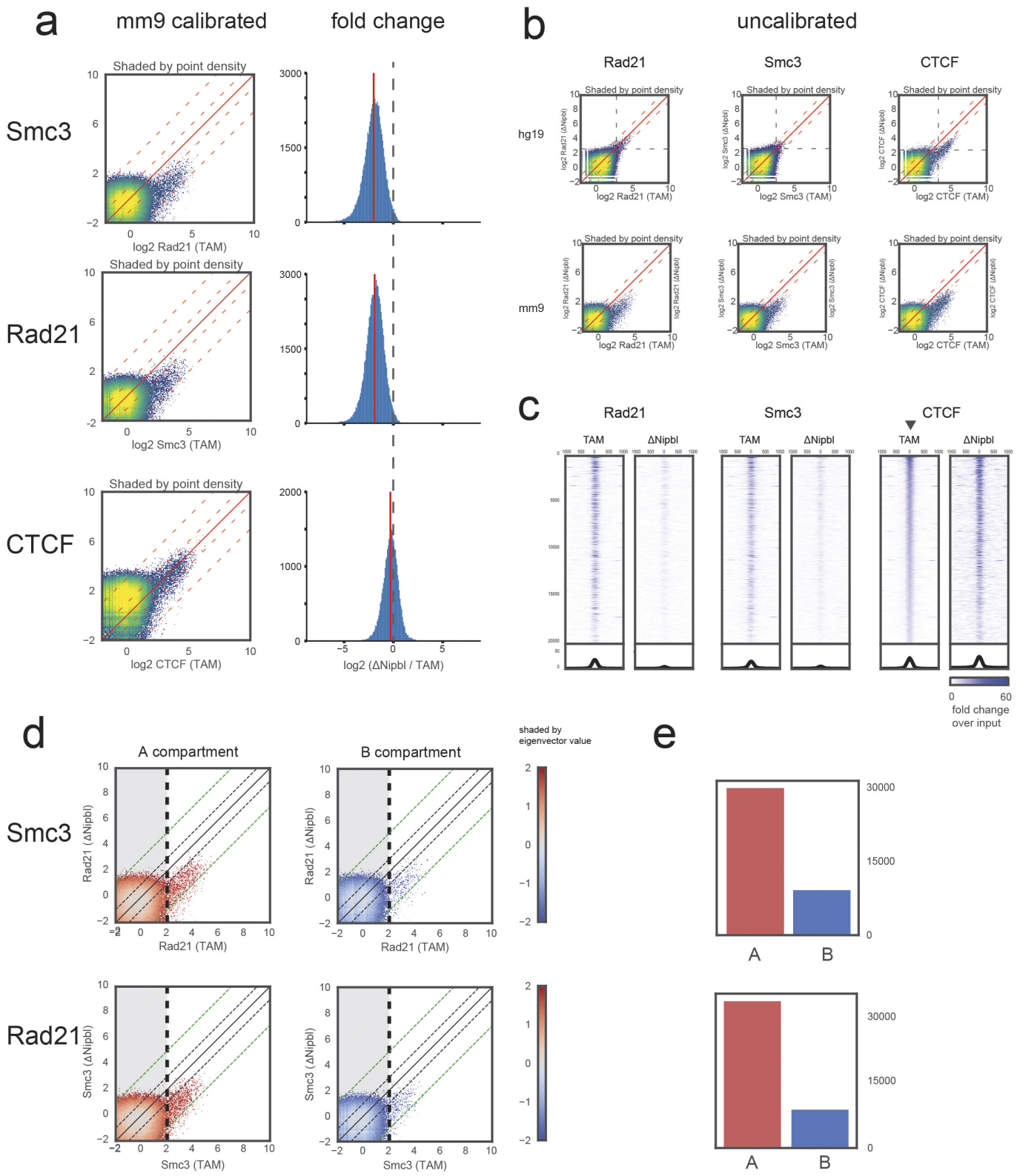
45. Tang, S.-H. E., Silva, F. J., Tsark, W. M. K. & Mann, J. R. A. A cre/*loxP*-deleter transgenic line in mouse strain 129S1/SvImJ. *Genesis* **32,** 199–202 (2002).
46. Logan, M. *et al.* Expression of Cre Recombinase in the developing mouse limb bud driven by a *Prxl* enhancer. *Genesis* **33,** 77–80 (2002).
47. Tannour-Louet, M., Porteu, A., Vaulont, S., Kahn, A. & Vasseur-Cognet, M. A tamoxifen-inducible chimeric Cre recombinase specifically effective in the fetal and adult mouse liver. *Hepatology* **35,** 1072–1081 (2002).
48. Li, W.-C., Ralphs, K. L. & Tosh, D. Isolation and culture of adult mouse hepatocytes. *Methods Mol. Biol.* **633,** 185–196 (2010).
49. Gonçalves, L. A., Vigário, A. M. & Penha-Gonçalves, C. Improved isolation of murine hepatocytes for *in vitro* malaria liver stage studies. *Malar. J.* **6,** 169 (2007).
50. Arrigoni, L. *et al.* Standardizing chromatin research: a simple and universal method for ChIP-seq. *Nuclei Acids. Res.* **44,** e67 (2016).
51. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9,** 999–1003 (2012).
52. Filippova, D., Patro, R., Duggal, G. & Kingsford, C. Identification of alternative topological domains in chromatin. *Algorithms Mol. Biol.* **9,** 14 (2014).
53. Girvan, M. & Newman, M. E. J. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99,** 7821–7826 (2002).
54. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer, 2006).
55. Bogu, G. K. *et al.* Chromatin and RNA maps reveal regulatory long noncoding RNAs in mouse. *Mol. Cell. Biol.* **36,** 809–819 (2015).
56. Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148,** 335–348 (2012).
57. Yue, F. *et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515,** 355–364 (2014).
58. Peric-Hupkes, D. *et al.* Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol. Cell* **38,** 603–613 (2010).
59. Harrow, J. *et al.* GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* **22,** 1760–1774 (2012).
60. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
61. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).
62. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
63. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods* **12,** 357–360 (2015).
64. Goloborodko, A., Imakaev, M. V., Marko, J. F. & Mirny, L. Compaction and segregation of sister chromatids via active loop extrusion. *eLife* **5,** e14864 (2016).
65. Eastman, P. *et al.* OpenMM 4: A reusable, extensible, hardware independent library for high performance molecular simulation. *J. Chem. Theory Comput.* **9,** 461–469 (2013).
66. Eastman, P. & Pande, V. S. Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *J. Comput. Chem.* **31,** 1268–1272 (2010).
67. Imakaev, M. V., Tchourine, K. M., Nechaev, S. K. & Mirny, L. A. Effects of topological constraints on globular polymers. *Soft Matter* **11,** 665–671 (2015).
68. Kawauchi, S. *et al.* Multiple organ system defects and transcriptional dysregulation in the *Nipbl*$^{+/-}$ mouse, a model of Cornelia de Lange Syndrome. *PLoS Genet.* **5,** e1000650 (2009).
69. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488,** 116–120 (2012).



Extended Data Figure 1 | Overview of various features of chromosomal architecture detected and quantified in Hi-C contact maps. Top row, intra-chromosomal maps show the decay of contact frequency with genomic distance, which can be quantified with the curves of contact frequency $P(s)$ versus genomic separation s. Middle row, both intra- and inter-chromosomal maps display a checkered pattern caused by compartmentalization of the genome. This pattern can be quantified by a continuous genomic track obtained via eigenvector decomposition of either *cis* or *trans* maps. Bottom row, intra-chromosomal maps at short genomic distance scales reveal domains of enriched contact frequency, which appear as bright squares along the main diagonal, and peaks which appear as bright dots connecting two loci. Both can be detected and quantified using specialized algorithms.

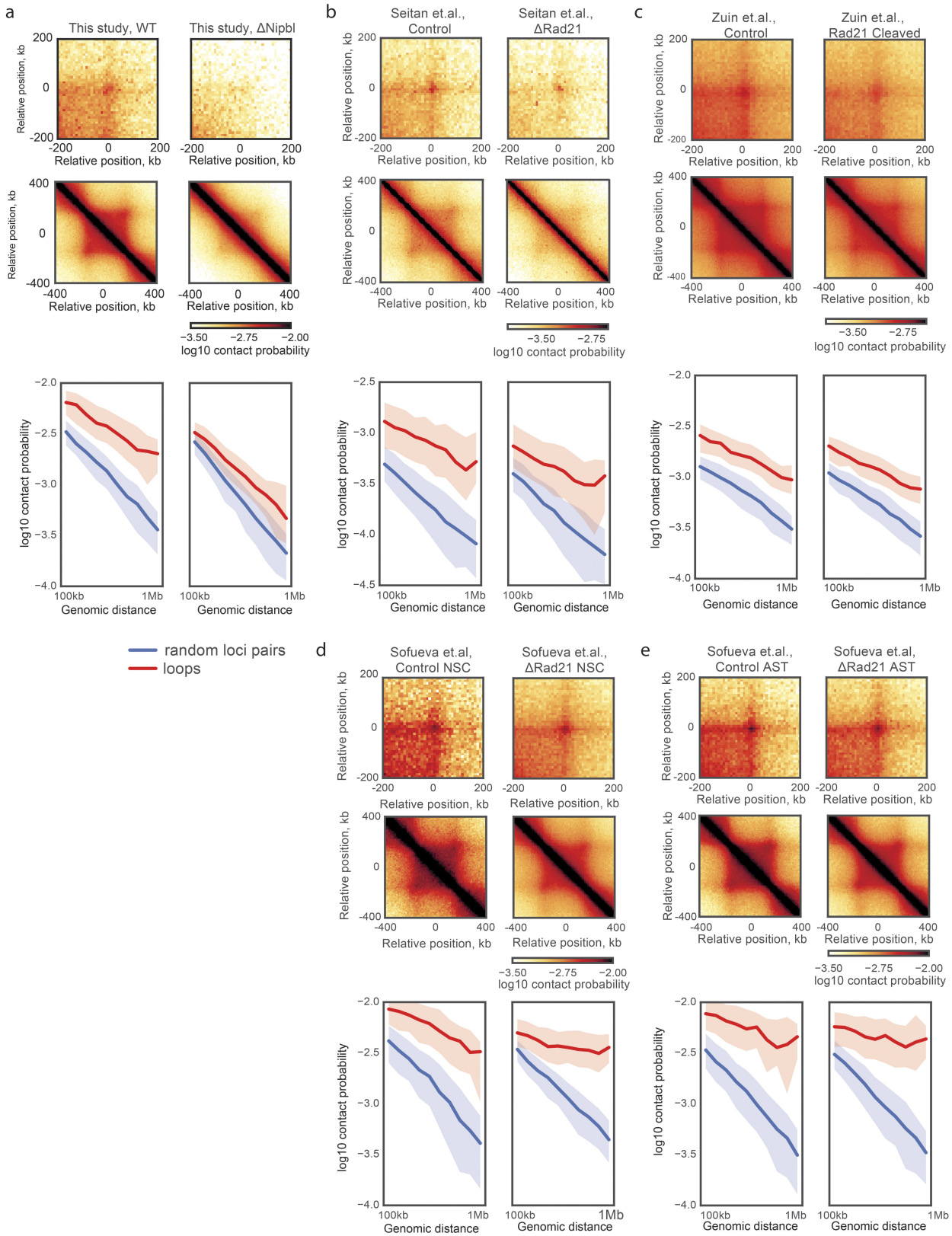


Extended Data Figure 2 | Conditional inactivation of *Nipbl* in mice. **a**, Schematic representation of the conditional allele, with *loxP* sites (red triangles) flanking exon 18. The reading frame of each exon is indicated below the corresponding square, as *x*–*x*. Deletion of exon 18 leads to a frame-shift introducing a premature stop codon (indicated by amino acids in red). The resulting protein lacks the critical HEAT domains that are conserved in NIPBL and SCC2 proteins. **b**, **c**, Embryonic day 12 (E12) embryos (**b**) and E18 fetuses (**c**) carrying the conditional *Nipbl* allele ($Nipbl^{flox}$) and either ubiquitous (Hprt:Cre[45]) or limb-specific (Prx1::Cre[46]) Cre recombinase drivers. Structures expressing Cre are rapidly lost in $Nipbl^{flox/flox}$ animals. Heterozygous $Nipbl^{flox/+}$ animals are grossly morphologically normal, but die soon after birth, as reported for other *Nipbl* loss of function alleles[68]. Fl, forelimb; md, mandible; abw, abdominal wall. **d**, **e**, Histochemical staining of liver section of adult Δ*Nipbl* mouse ($Nipbl^{flox/flox}$; Ttr::CreERT2; 10 days after tamoxifen injection) for a proliferation marker (Phos-H3) (**d**) and apoptosis (cleaved Caspase3) (**e**) (both shown in red). Nuclei are stained with DAPI (blue). Staining was performed once.

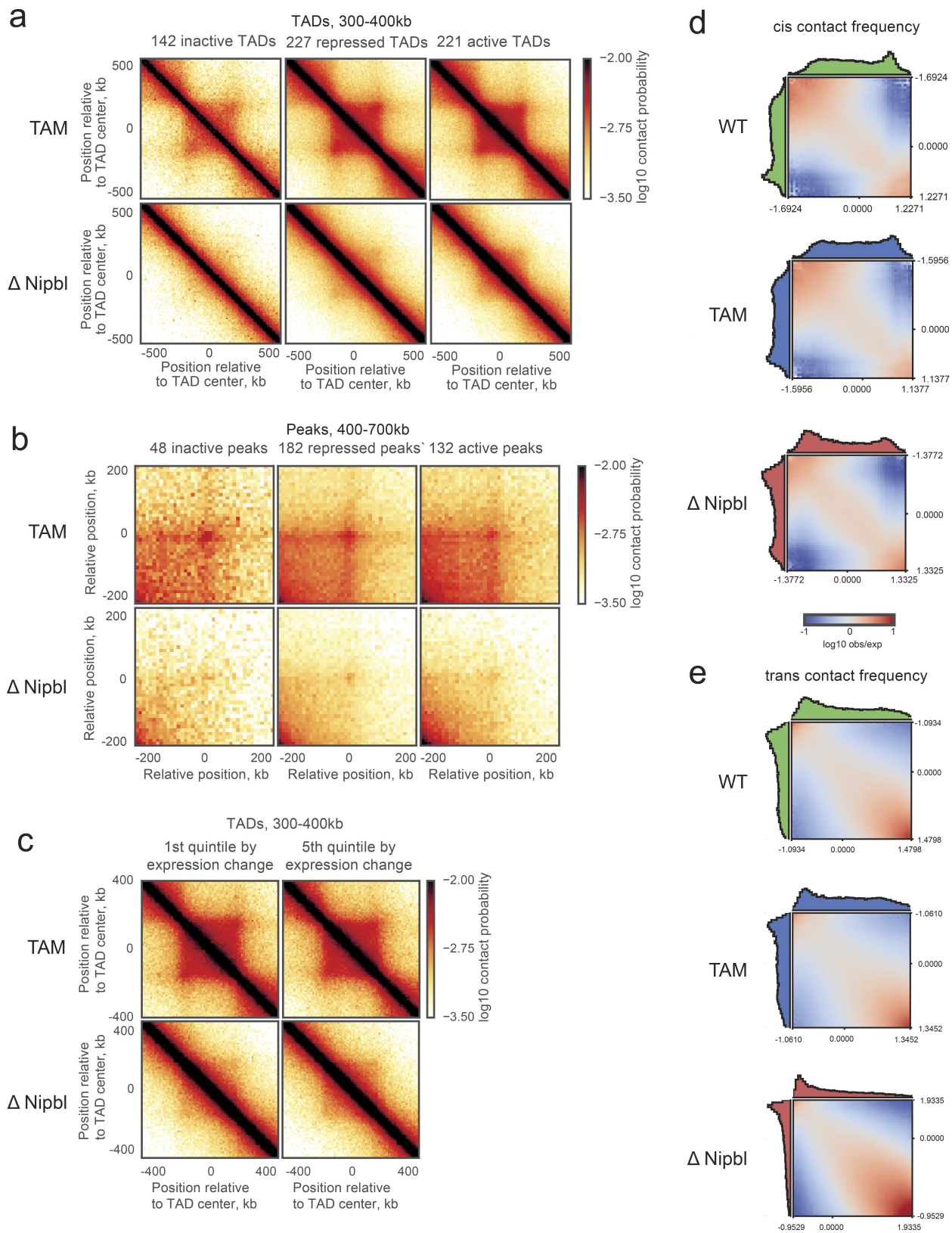


Extended Data Figure 3 | See next page for caption.

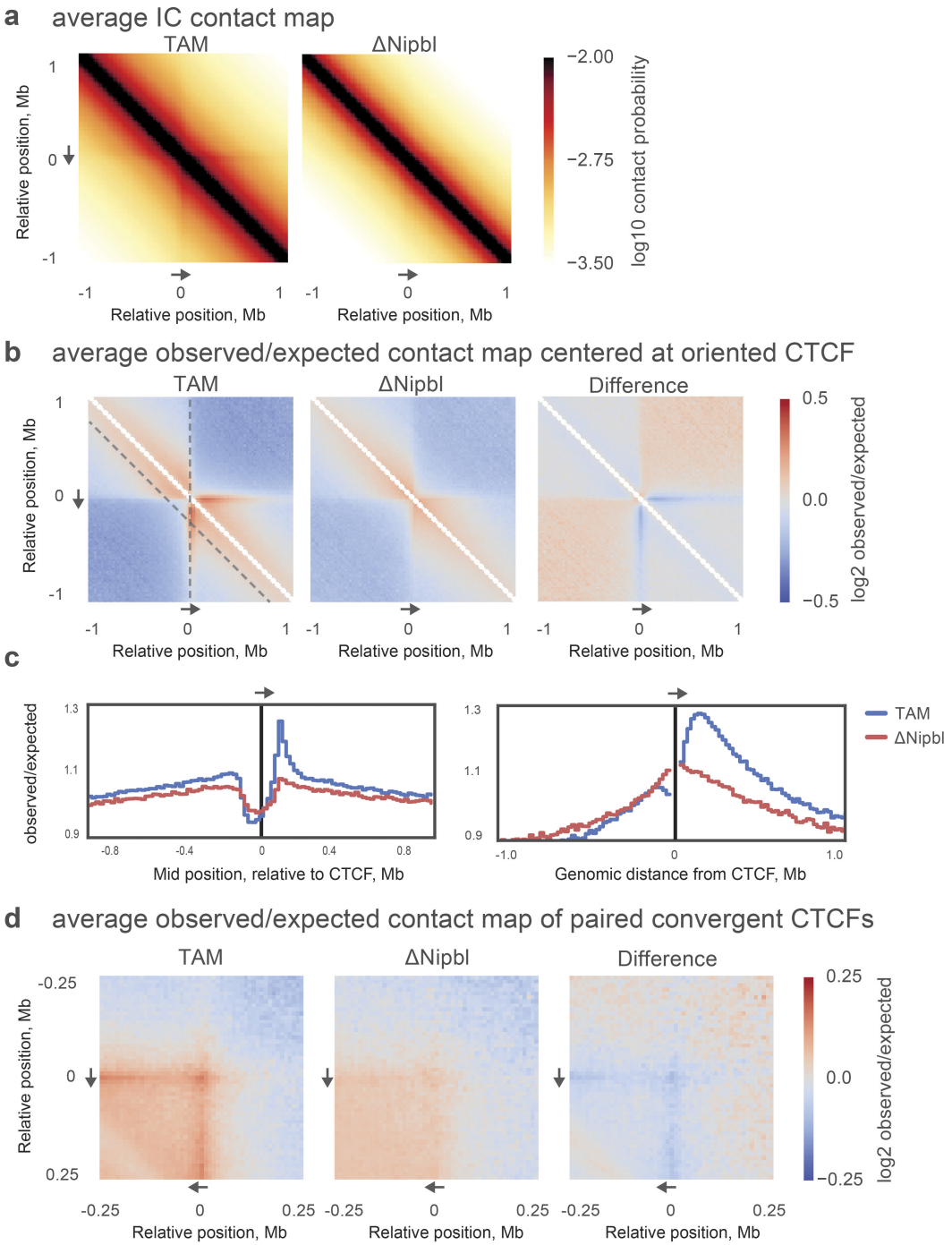
Extended Data Figure 3 | Calibrated ChIP–seq for CTCF and cohesin (Rad21 and SMC3). a, Left, comparison of calibrated mouse ChIP–seq data. For each factor (CTCF, Rad21, SMC3), we used the ratio of the top 0.2% of 200-bp bins (~29,000 points) in the TAM hg19 fraction vs the Δ*Nipbl* hg19 fraction to rescale the mm9 signal in the Δ*Nipbl* condition in order to be compared with the mm9 signal in the TAM condition (see Methods). The scatter plot heatmaps in the left column are shaded by point density. Right, $\log_2$ fold difference ratio between Δ*Nipbl* and TAM for ChIP–seq peaks (points with ChIP in TAM > 2.0). While the calibrated CTCF binding signal stays relatively constant between conditions (within the twofold envelope), most of the calibrated cohesin binding signal drops by about 2–8-fold in Δ*Nipbl* cells, with a mean depletion of 3.7-fold. **b**, Uncalibrated ChIP–seq reads used for the calibration of human cell (hg19, left) and mouse (mm9, right). Processed mapped reads were converted to genome-wide signal tracks binned at 200-bp resolution. Scatter plots of these genomic tracks (TAM vs Δ*Nipbl*) are shown as heatmaps shaded by point density. For cohesin subunits, the hg19 signal has a similar profile in both conditions, but the Δ*Nipbl* signal is diminished in the uncalibrated mm9 fraction. The uncalibrated mm9 signal also appears to be reduced in Δ*Nipbl* for CTCF; however, a similar diminishing effect is seen in the hg19 data. **c**, Stacked heatmaps of calibrated ChIP–seq signal at the top 20,000 CTCF binding sites (peaks with an assigned CTCF motif), ranked by fold change over input in the TAM control condition. **d**, Scatter plots of calibrated ChIP–seq tracks as in **b**, but split into two groups by compartment type A (compartment eigenvector >0) and B (eigenvector assignment <0). The shading is coloured by the eigenvector signal. Black and green dashed diagonal lines demarcate the twofold and eightfold envelopes, respectively. **e**, Total cohesin occupancy. Bar plot showing the number of ChIP–seq 200-bp points in the non-shaded area of the scatter plots (bins with TAM signal >2) representing high confidence binding signal in the wild type. While cohesin binding is more than threefold more prevalent in A-compartment regions, the scatter plots show that both A and B regions respond equally to *Nipbl* deletion.



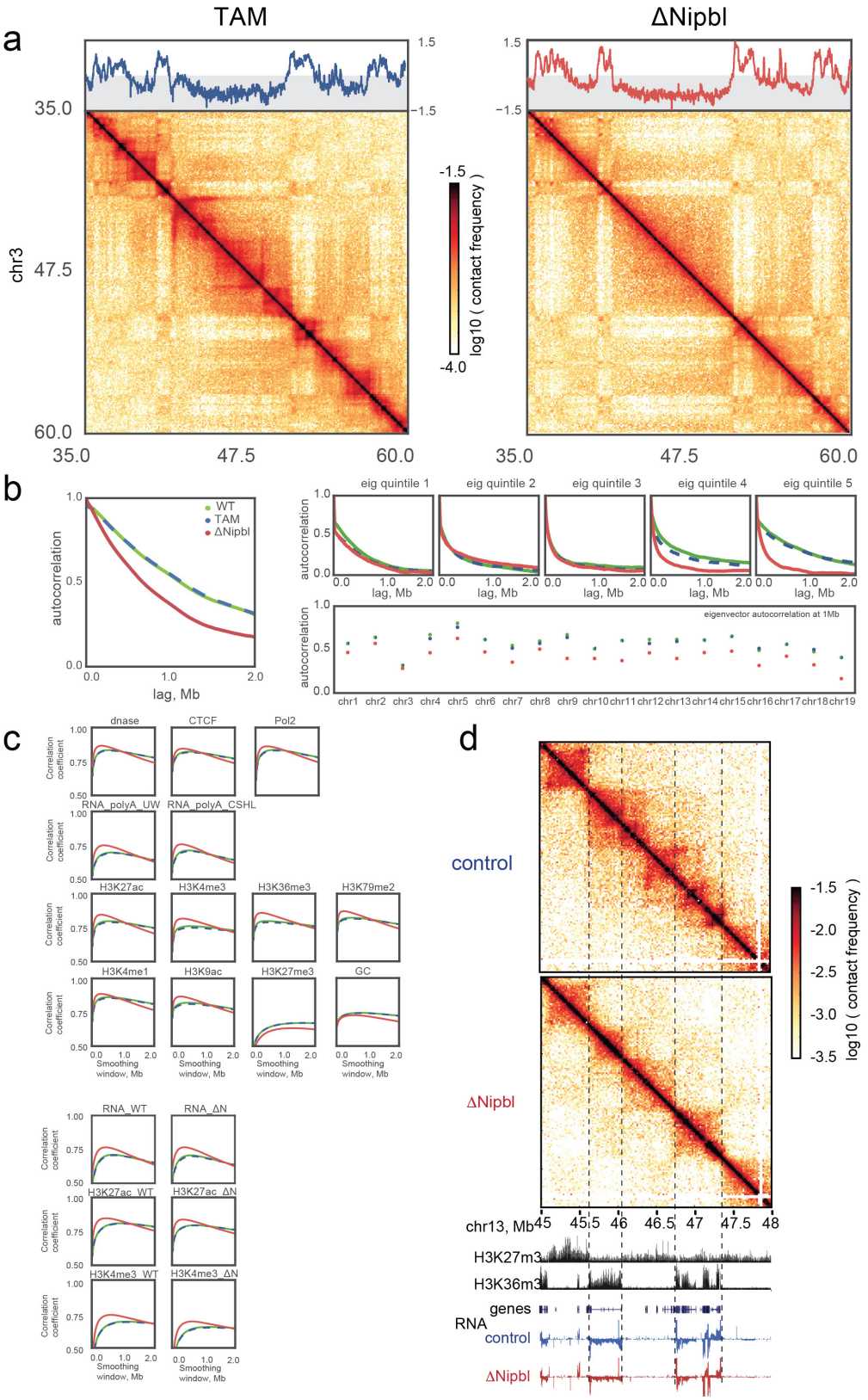
Extended Data Figure 4 | Deletion of *Nipbl* in this study leads to more robust disappearance of TADs and associated peaks than in previous publications. a, Genetic deletion of *Nipbl* in hepatocytes (this study). **b**, Deletion of *Rad21* in thymocytes[20]. **c**, Proteolytic cleavage of RAD21 in HEK293T cells[19]. **d**, **e**, Deletion of *Rad21* in NSCs and astrocytes (ASTs)[21]. For each data set: top, average Hi-C map around 102 peaks with size range 500–600 kb[12] in wild-type and Δ*Nipbl* contact maps; middle, average Hi-C map of TADs called in each data set; bottom, relative contact probability between pairs of peak loci versus genomic distance, compared to randomly selected pairs of loci. The thick line shows the median contact probability; the shading shows the envelope between the 25th and 75th percentiles of contact probability at each genomic separation. Note substantial change in the contact probability at peaks (red line) upon *Nipbl* depletion (**a**) and little change in other studies (**b**–**e**). All studies used comparable Hi-C sequencing depth (Supplementary Table 3).



Extended Data Figure 5 | Residual structures are observed in active and repressed regions of the genome after *Nipbl* deletion. For each TAD, an activity was assigned based on the dominant simplified ChromHMM state category. **a**, Average contact map of 300–400 kb TADs in inert, repressed and active regions of the genome. **b**, Average contact map of 300–700 kb peaks in inert, repressed and active regions of the genome. **c**, Average contact maps of most upregulated 20% (left) and most downregulated 20% (right) of 300–400 kb TADs. **d**, **e** Compartmentalization saddle plots in *cis* (**d**) and *trans* (**e**): average interaction frequencies between pairs of loci (100-kb bins) arranged by their compartment signal (eigenvector value). The interaction frequencies in *cis* (top row) are computed for observed and expected contact maps. Notice enrichment of AA and depletion of AB interactions in Δ*Nipbl* cells. Histograms along the axes show the distributions of eigenvector values.

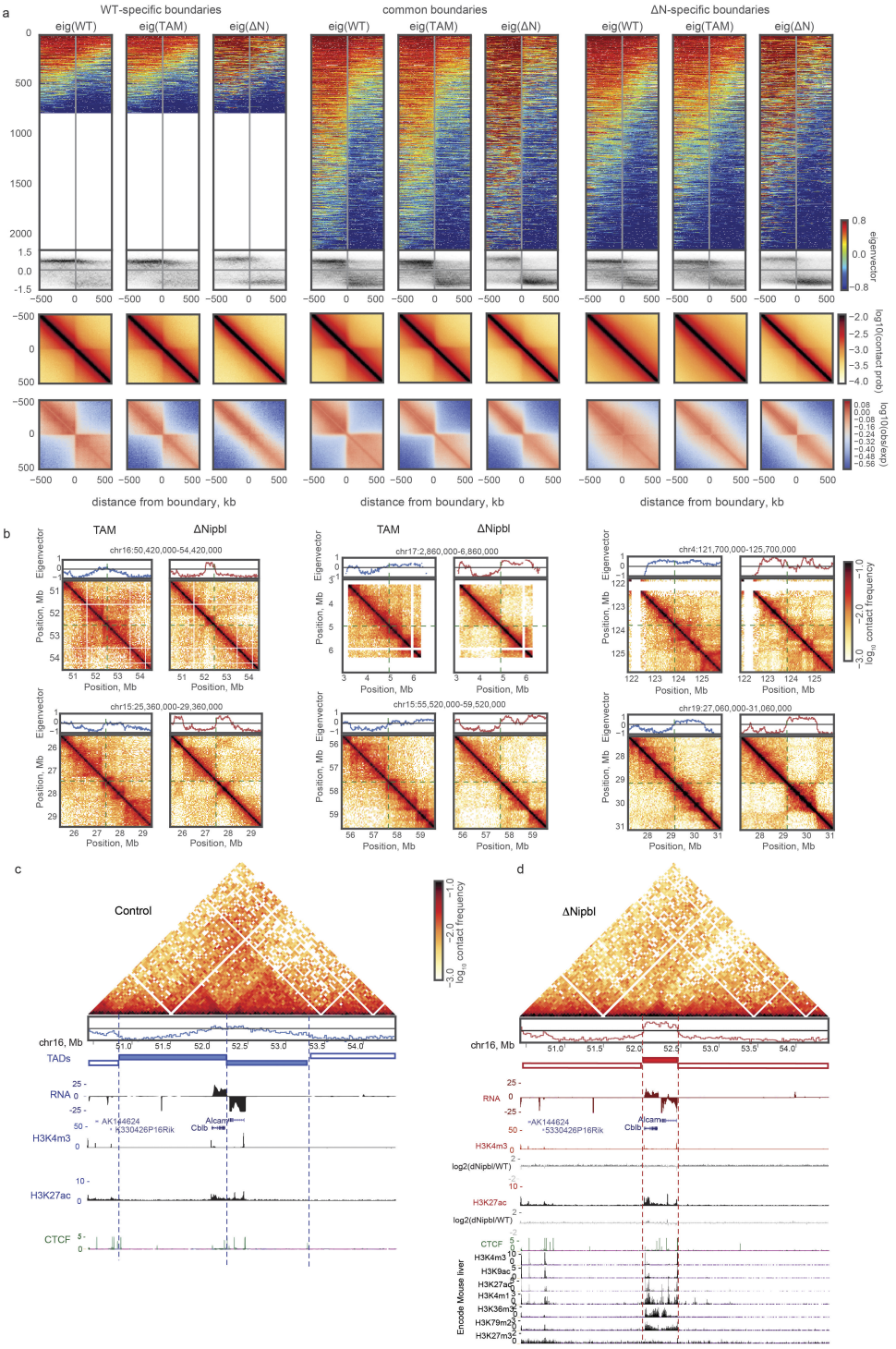


Extended Data Figure 6 | Average Hi-C contact footprint of CTCF sites. CTCF peaks were detected in our ChIP–seq data from TAM control cells and supported by an underlying CTCF binding motif occurrence. **a**, Average iteratively corrected 20-kb resolution contact map of around 42,000 CTCF peaks in TAM and Δ*Nipbl* cells. Individual contributing snippets to the composite heatmap were oriented such that the CTCF motif points in the direction shown by the grey arrow. **b**, Average 20-kb resolution contact map normalized by the expected contact frequency at a given genomic separation. **c**, Average observed over expected contact frequency curves along 'slices' of the composite heatmaps depicted by dashed grey lines in **b**: left, insulation profile at 200 kb separation (diagonal dashed line in **b**); right, 'virtual 4C' curve (vertical dashed line in **b**) of the composite heatmap. **d**, Average 10-kb resolution observed over expected contact map centred on about 11,000 pairs of CTCF ChIP–seq peaks with convergent motif orientations separated by 200 ± 10 kb.



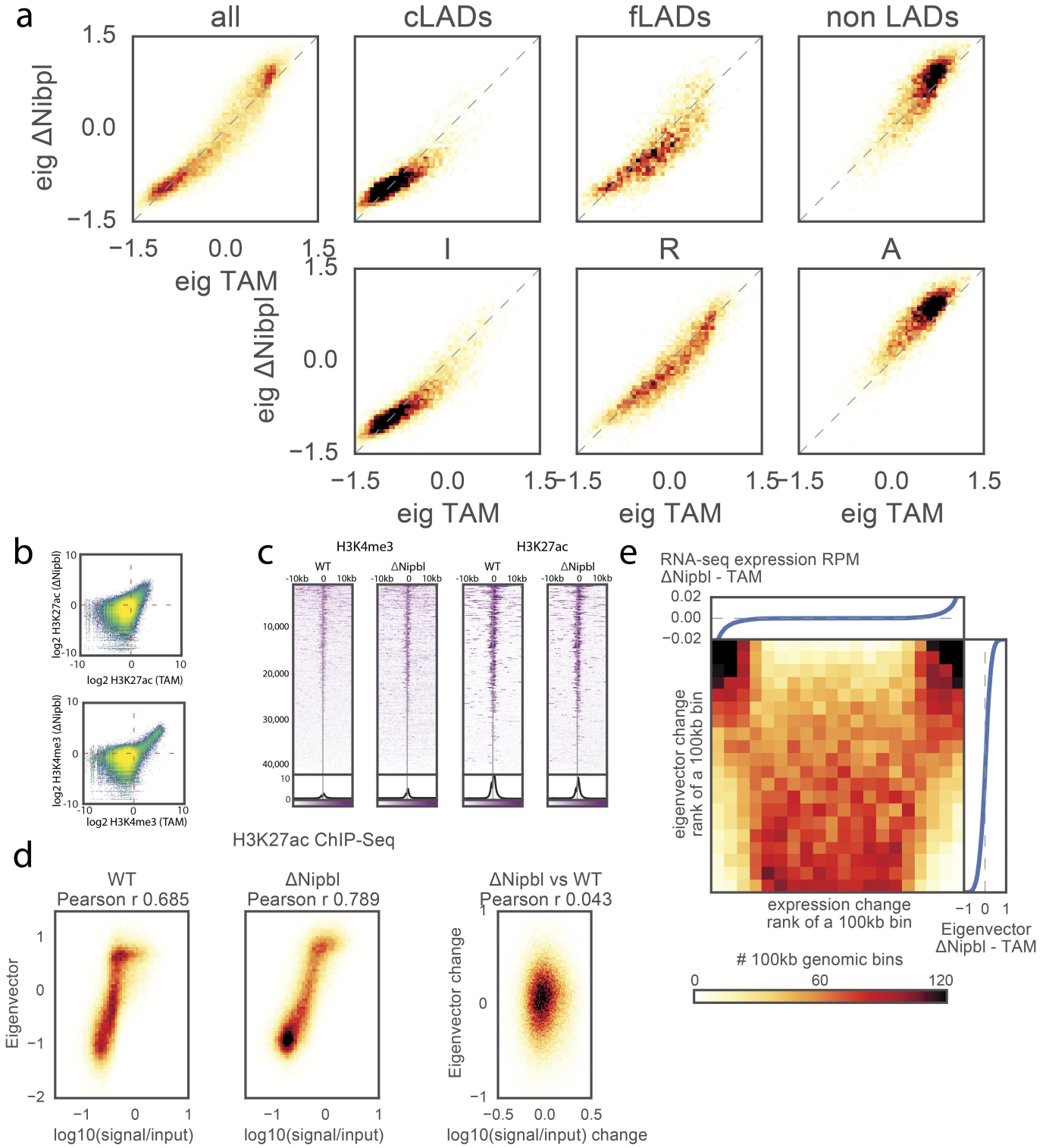
Extended Data Figure 7 | See next page for caption.

Extended Data Figure 7 | Fragmentation of A and B compartment-type in smaller alternating regions upon *Nipbl* deletion is activity-dependent. a, Example region (chr3: 35–60 Mb) illustrating lack of compartment fragmentation in predominantly B-type regions with robust disappearance of TADs. Top, compartment eigenvector; bottom, contact matrix snapshot. **b**, Autocorrelation of eigenvector tracks reveals genome-wide fragmentation of active compartments. Left, genome-wide Spearman coefficient of correlation of the 20-kb *cis* eigenvector values ($n = 113{,}372$) of pairs of loci as a function of their genomic separation (autocorrelation). Top right, eigenvector correlation of locus pairs split by quintile of the eigenvector value of the upstream locus. Bottom right, chromosome-wide values of eigenvector correlation of locus pairs separated by 1 Mb. **c**, Spearman coefficient of correlation between the smoothed histone ChIP–seq, transcription factor ChIP–seq and RNA-seq tracks and the 20-kb *cis* eigenvectors ($n = 113{,}372$) as a function of the smoothing window size. Left group of panels show ENCODE data, right show data from this study. First and second rows, histone marks; third row, RNA-seq tracks; fourth row, miscellaneous tracks (DNase hypersensitivity, CTCF and PolII ChIP–seq and GC content). Δ*Nipbl* eigenvectors show an increased correlation with tracks associated with transcriptional activity but a decreased correlation with the repression-associated track of H3K27me3 and GC content. **d**, Example of a large continuous wild-type (top) A-type compartment region (chr13: 45–48 Mb) and the same region in Δ*Nipbl* (bottom). Δ*Nipbl* compartment transitions highlighted by black dashed lines. TAD boundaries in the wild type are shifted or lost and replaced by compartment transitions in Δ*Nipbl* cells. Below, histone ChIP–seq tracks[69] and stranded RNA-seq tracks (blue, TAM hepatocytes; red, Δ*Nipbl* cells) highlight that wild-type and TAM TADs do not strictly follow the underlying chromatin activities, whereas the new checkered pattern in Δ*Nipbl* cells delineated by dashed lines correspond precisely to active versus inactive chromatin domains. In **a** and **d**, both replicates of each condition show similar results.

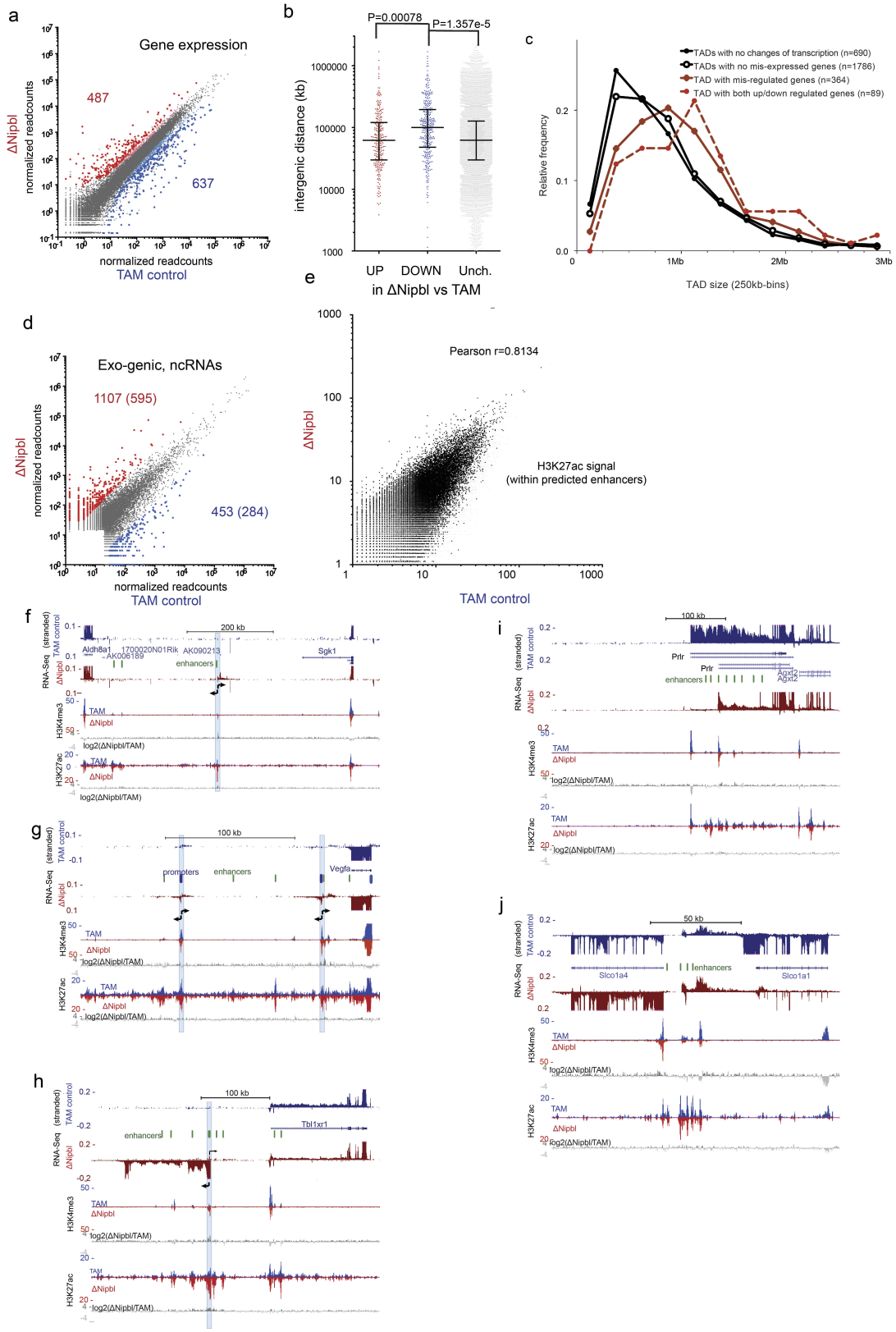


Extended Data Figure 8 | See next page for caption.

Extended Data Figure 8 | TADs and compartments constitute independent layers of genome organization. a, The residual contact-insulating boundaries in Δ*Nipbl* cells are associated with compartment transitions. The first group of columns considers the boundaries detected in wild-type cells only, the second pair considers boundaries detected in both wild-type and Δ*Nipbl* cells, and the last pair considers boundaries detected in Δ*Nipbl* cells only. For each group, the first, second and third columns display data (eigenvectors and Hi-C) from wild-type, TAM and Δ*Nipbl* cells, respectively. Within each column the top row is a stack of eigenvector tracks in a ±500-kb window around boundaries, oriented such that the left half of the window has greater average signal value and sorted by the average wild-type eigenvector value in the window. The second row shows a density histogram of eigenvector values as a function of the distance to the boundary. The third and fourth rows show the boundary-centred average contact probability and observed/expected contact ratio, respectively. The density histograms show that common and Δ*Nipbl*-specific boundaries correspond to sharp transitions of compartment signals in Δ*Nipbl* cells, in contrast to the more diffuse signal at these positions in wild-type and TAM cells. **b**, Boundaries of former TADs and new compartment domains do not coincide. Examples of TADs detected in wild-type cells, which contain sharp compartment transitions revealed by Δ*Nipbl* contact maps. Left, TAM control data; right, Δ*Nipbl* data. Top of each figure, local compartment signal in the corresponding cell type. The contact maps are centred at the sharp compartment transition in Δ*Nipbl*. These examples illustrate that that chromatin-bound cohesins can interfere locally with genome compartmentalization. **c**, **d**, New compartments do not respect TAD boundaries but do respect the underlying chromatin domains. A large region (chr16: 50,420,000–54,420,000) adopts a very different 3D organization in control (**c**, in blue) and Δ*Nipbl* cells (**d**, in red). Hi-C data are shown, as well as the eigenvector values in the two conditions. RNA-seq tracks showed minimal changes of expression (*Alcam* expression is reduced twofold in Δ*Nipbl* cells) and chromatin states. ChIP–seq tracks for H3K27ac and H3K4me3 are shown in the two conditions, with $\log_2$ ratio tracks under the Δ*Nipbl* (**d**) panel. Encode tracks (corresponding to wild-type liver cells) are shown in the grey boxed area. The new structure adopted in Δ*Nipbl* cells put together the two active genes that are normally in different TADs in the same domain, corresponding to the active chromatin linear domain. In **b**–**d**, both replicates of each condition showed similar results.



Extended Data Figure 9 | Eigenvector change upon *Nipbl* deletion is activity-dependent and uncorrelated with changes in gene expression or epigenetic marks. a, Correlation of *cis* eigenvector values of 100-kb genomic bins before and after *Nipbl* deletion, split by the functional state of chromatin. Top row, left to right: genome-wide relationship; bins showing constitutive lamin-B1 association across four mouse cell types (cLADs); bins showing variable (facultative) lamin-B1 association (fLADs); binds not showing any association (non LADs). Bottom row: bins assigned the inert ChromHMM simplified state; bins assigned the repressed state; bins assigned the active state. **b**, Scatter plot of genome-wide ChIP–seq signal binned at 200 bp in wild-type versus Δ*Nipbl* cells. Top, H3K27ac; bottom, H3K4me3. **c**, Stacked heatmap of histone ChIP–seq signal over input ± 10 kb around putative TSS sites sorted by total H3K27ac signal in the wild type and oriented by TSS strand. From left to right: H3K4me3 in wild-type and Δ*Nipbl*, followed by H3K27ac in wild-type and Δ*Nipbl*. **d**, ChIP–seq signal for histone marks of activity versus eigenvector value of 20-kb bins. Top row, H3K27ac; bottom row, H3K4me3. Left column, wild-type cells; middle column, Δ*Nipbl* cells; right column, correlation of changes in both signals upon *Nipbl* deletion. **e**, Change in compartment structure upon *Nipbl* deletion cannot be attributed to the sign of the local expression change. The heatmap shows the number of 100-kb genomic bins as a function of the ranks of expression change and the eigenvector change. The attached plots show the correspondence between the values of expression change (top) or eigenvector change (right) and their ranks.



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | Expression changes in Δ*Nipbl* hepatocytes. **a**, Changes in gene expression between TAM controls and Δ*Nipbl* liver cells (four replicates for each condition) analysed with DESeq2[62]. Genes with significant changes in gene expression (FDR >0.05) are coloured red (upregulated, $n = 487$) or blue (downregulated, $n = 637$), with larger dots corresponding to genes with a fold-change of more than 3. **b**, Intergenic distances for the different categories of disregulated genes (with fold-change more than 3; upregulated, 268; downregulated, 350; unchanged, 15,055). Statistical differences determined by an unpaired two-tailed *t*-test. The differences between means were 50,020.40 (CI 95% = 27723.85–72316.95) and 52,185 (CI 95% = 21824–82547) for downregulated versus unchanged, and downregulated versus upregulated, respectively. **c**, Size distribution of the TADs observed in wild-type cells (lost in Δ*Nipbl* cells) depending on the degree alteration of their transcriptional states. The size of TADs with transcriptional changes (red) is significantly larger than those that do not show transcriptional alterations (black) (Kolmogorov–Smirnov, $P = 4.095 \times 10^{-8}$). **d**, Change in transcription in non-genic intervals (including intergenic and antisense within gene bodies). Gene expression was calculated as the normalized number of reads within intervals defined by merging adjacent 1-kb windows showing readcounts over background (see Methods). The numbers of non-coding transcripts upregulated (in red) or downregulated (in blue) in Δ*Nipbl* compared to the TAM control is given ($P < 0.01$ using a two-tailed *t*-test, four replicates per condition, fold-change higher than 8), with the second number indicating the high-confidence events (labelled with coloured dots, expression value over an arbitrary threshold of 30 reads) which constitute the list used for subsequent analyses. **e**, Comparison of control and Δ*Nipbl* H3K27ac normalized signals within predicted liver enhancer elements ($n = 51,850$; readcounts within ±500 bp of predicted enhancer peak)[69]. **f**–**i**, Examples of transcriptional changes upon *Nipbl* deletion. Stranded RNA-seq and ChIP–seq tracks (H3K4me3, H3K27ac) are shown for control (blue) and Δ*Nipbl* (red) samples. Comparison of the chromatin profiles are shown with $\log_2$[Δ*Nipbl*/TAM] tracks for H3K4me3 and H3K27ac (in grey). Active enhancers (peaks of high H3K27ac, H3K4me1, low H3K4me3)[69] are depicted as green ovals. **f**, Chr10: 21,090,000–21,781,000. Bidirectional transcription (position labelled with a blue bar) arises from an isolated enhancer in Δ*Nipbl* cells. **g**, Chr17: 45,945,000–46,176000. Bidirectional transcription (position labelled with a blue bar) arises from two cryptic promoters (H3K4me3 peaks, no or weak transcription in TAM) downstream of *Vegfa*. **h**, Chr3: 21,712,500–22,126,240. A new transcript from a cryptic promoter 100 kb upstream of *Tbl1xr1*. H3K27ac signal is enhanced at peaks surrounding the activated cryptic promoter. **i**, Chr15: 9,873,000–10,354,700. Promoter switch for *Prlr*, from an upstream promoter to a more downstream one surrounded by active enhancers. **j**, chr6: 141,743,961–141,904,692. Downregulation of *Slco1a1* and concomitant upregulation of *Slco1a4* and noncoding intergenic transcripts (arrowheads). Distance of *Slco1a4* promoter to intergenic enhancers is less than 10 kb, compared to 80 kb for *Slco1a1*.

nature research

Corresponding author(s): Francois Spitz, Leonid Mirny

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Sample size was not pre-determined, We used sample sizes commonly used and accepted for the type of experiments. For animal experiments, we used at least 3 animals per group (range 3 to 17) to allow basic statistical inference while using a justifiable number of mutant mice. We used 4 biological replicates per conditions for RNA-Seq, 2 for Hi-C, and 2 for ChiP-Seq. Hi-C data was subsequently pooled. For ChIP-Seq, representative tracks from one of the replicates are shown.

2. Data exclusions

   Describe any data exclusions.

   No data were excluded from the analysis, except for Hi-C. As recommended for the analysis of Hi-C data, we removed bins with low-coverage as well as contact signals between same of adjacent genomic bins.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   all replication attempts were successful

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   allocation of mice to one or the other group was not randomized. However, samples were treated in parallel, and whenever possible with internal controls, so as to equally match potential confounding effects

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   Experiments were not blinded

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	Statistical analyses were performed with R (R studio) and Prism 6. Genomic analyses used Galaxy and ENCODE tools, Hiclib, STAR V2.5.0a, HTSeq, DESeq2, HISAT2. Details are included in the Method section. Custom codes used are described in Method section and Code availability section.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique material have been used. Mouse strains - as cryopreserved sperm - are available for distribution by the authors, upon request.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	αH3k4me3: C15410003-50, Diagenode (valided by the manufacturer) αH3K27Ac: ab4729, Abcam; lot GR7675-1 (valided by the manufacturer) Anti-Rad21: Abcam ab992; lot GR214359-7 (valided by the manufacturer) Anti-CTCF: Millipore 07-729; lot 2887267 (valided by the manufacturer) anti-SMC3: Abcam ab9263, lot GR290533-3 (valided by the manufacturer) anti SMC1: provided by Ana Losada (described/validated in 10.1038/emboj.2012.11) anti SA1: provided by Ana Losada (described/validated in 10.1038/emboj.2012.11)

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	HEK293 cells (originally from the Trono lab - EPFL - Lausanne - Switzerland)
b. Describe the method of cell line authentication used.	HEK293 cells were not authenticated
c. Report whether the cell lines were tested for mycoplasma contamination.	HEK293 cells were not tested for mycoplasma
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	cell line used (HEK-293) is not in the ICLAC database

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	we used animals from the Nipbl [flox] allele made in the lab (internal reference clone 3H6), crossed with Ttr-cre/Esr1 strain (MGI:3046546 Tg(Ttr-cre/Esr1*)1Vco). Mice used were adult (10-20 week-old, usually 12 week-old) of both sexes, except for Hi-C experiments for which only males were used.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	the study did not involve human participants

nature research

Corresponding author(s): Francois Spitz, Leonid Mirny

☐ Initial submission ☐ Revised version ☒ Final submission

ChIP-seq Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data deposition

1. For all ChIP-seq data:

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

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

2. Provide all necessary reviewer access links.
The entry may remain private before publication.

https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE93431

WT vs TAM vs ΔNipbl, locked by zoom and location
• http://higlass.io/app/?config=fLYyoNpETSe-QbXg50_DOA
Same, with overlaid eigenvector tracks:
• http://higlass.io/app/?config=RNj34_97T3SWJ5RcSDonjQ
WT vs ΔNipbl with zoom-ins connected by view projections (grey squares)
• http://higlass.io/app/?config=Tf2-ublRTey9hiBKMlgzwg

UCSC tracks (ChIP-Seq and RNA-Seq)
http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=spitzfr&hgS_otherUserSessionName=Ko_Nipbl_Public0

3. Provide a list of all files available in the database submission.

4. If available, provide a link to an anonymized genome browser session (e.g. UCSC).

• http://higlass.io/app/?config=fLYyoNpETSe-QbXg50_DOA
• http://higlass.io/app/?config=RNj34_97T3SWJ5RcSDonjQ
• http://higlass.io/app/?config=Tf2-ublRTey9hiBKMlgzwg

UCSC tracks (ChIP-Seq and RNA-Seq)
http://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=spitzfr&hgS_otherUserSessionName=Ko_Nipbl_Public0

Methodological details

5. Describe the experimental replicates.

See Method section.

6. Describe the sequencing depth for each experiment.

Condition Total Mapped Unique
Nipbl_8094_ChIPK27ac 28888231 28138779 22336333
Nipbl_8094_ChIPK4me3 19322681 18765579 13559526
Nipbl_8094_Input 20708221 20033518 14857648
WT_8097_ChIPK27ac 23893468 23425426 18952812
WT_8097_ChIPK4me3 21546139 21033793 15429627
WT_8097_Input 22718479 22040698 17235134
Rad21Ctrl 55861956 52702007 39862949
Rad21Delta 50724880 48735141 36299305
Smc3Ctrl 53349512 50898271 38132996
Smc3Delta 59299518 56890933 43428069
InputCtrl 55442998 53763091 36651837

InputDelta 64217332 60983192 43185393
CtcfCtrl 33078713 32168242 22805555
CtcfDelta 43884331 42740092 32374478

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

αH3k4me3: C15410003-50, Diagenode
αH3K27Ac: ab4729, Abcam; lot GR7675-1
Anti-Rad21: Abcam ab992; lot GR214359-7
Anti-CTCF: Millipore 07-729; lot 2887267
anti-SMC3: Abcam ab9263, lot GR290533-3

all those antibodies have been used for ChIP-Seq in multiple publications

8. Describe the peak calling parameters.

We processed ChIP-seq data following the steps of the ENCODE pipeline [https://github.com/ENCODE-DCC/chip-seq-pipeline]. Alignment: we used bwa 0.7.12 (program bwa aln with parameters: -q 5 -l 32 -k 2). Filtering: PCR duplicates were marked using picardtools 2.7.1. Unmapped reads, non-primary alignments, and low quality alignments (mapQ < 30) were removed using samtools 1.3. Cross-correlation analysis was performed using phantompeakqualtools. Peaks and signal tracks were generated using MACS2. Peak calls were generated using MACS2 with parameters --p 1e-2, --nomodel, --shift 0, --keep-dup all. For Rad21, Smc3 and CTCF ChIP-seq, we followed the same steps with the following alterations: reads from pooled mouse hepatocyte chromatin and HEK human chromatin (internal control and calibration) were mapped to the combined mm9 and hg19 assemblies using the bwa mem program with default parameters. After filtering, reads were divided into those that mapped uniquely to either mm9 or hg19.

9. Describe the methods used to ensure data quality.

Raw read files were assessed using fastqc prior to processing. Cross-correlation and phantom peak analysis was used to ensure data quality and estimate fragment lengths. All datasets used for peak calling received a quality tag of 1 (High) or 2 (veryHigh) from cross-correlation analysis using phantompeakqualtools. Peak statistics are provided in the following table.
name treatment control total_peaks FDR_5_percent gt_5_fold_fc
CtcfCtrl CtcfCtrl InputCtrl 71347 53920 32488
CtcfDelta CtcfDelta InputDelta 43277 23347 9805
Rad21Ctrl Rad21Ctrl InputCtrl 44306 37094 20782
Rad21Delta Rad21Delta InputDelta 20730 7216 2577
Smc3Ctrl Smc3Ctrl InputCtrl 44306 37094 20782
Smc3Delta Smc3Delta InputDelta 31368 9683 2457

10. Describe the software used to collect and analyze the ChIP-seq data.

We followed and used the same software tools used in the ENCODE ChIP-seq pipeline (https://github.com/ENCODE-DCC/chip-seq-pipeline): bwa, picard, phantompeakqualtools, MACS2, bedtools, kent UCSC binaries.