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 Haering,[9] Leonid Mirny,[3,5,#] and Francois Spitz[1,4,6,7,#]

Imaging and chromosome conformation capture studies have revealed several layers of chromosome organization, including segregation into megabase-large active and inactive compartments, and partitioning into sub-megabase domains (TADs). Yet, it remains unclear how these layers of organization form, interact with one another and impact genome functions. Here, we show that deletion of the cohesin-loading factor *Nipbl*, in mouse liver, leads to a dramatic reorganization of chromosomal folding. TADs and associated peaks vanish globally, even in the absence of transcriptional changes. In 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 3D organization of the genome results from the interplay of two independent mechanisms: 1) cohesin-independent segregation of the genome into fine-scale compartments, defined by chromatin state; 2) cohesin-dependent formation of TADs, possibly by loop extrusion, which contributes to guide distant enhancers to their target genes.

The three-dimensional organization of chromosomes is tightly related to their biological function[1–3]. Hi-C maps have revealed key features of the 3D organization of metazoan chromosomes, including compartmentalization, TADs, and interaction peaks[4,5](Extended Data Figure 1).

Compartmentalization is visible as a characteristic checkerboard pattern of contact enrichment both in *cis* and *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 s harp 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 corner of TAD squares. Despite a lack of supporting mechanistic experiments, compartments, TADs and peaks are typically assumed to constitute hierarchical levels of chromosome folding. Yet, the connection between them remains poorly understood.

Architectural proteins, notably cohesin and CTCF, are believed to play crucial roles for interphase chromatin organization[13]. Cohesin and CTCF co-localize at TAD boundaries and the bases of Hi-C peaks[7,12], but their roles are not fully defined. The recently proposed loop extrusion model[14,15] yields predictions that are consistent with experiments. In this model, TADs emerge from the progressive extrusion of chromatin loops by a protein complex (e.g. cohesin) until it dissociates from chromatin or reaches a boundary element (e.g. CTCF sites). Recent experimental manipulation of CTCF motifs[15–17] and

CTCF depletion[18] support the proposed role of CTCFs as boundary elements, but cohesin's role in interphase chromatin organization and the process of loop extrusion remains ambiguous, as experimental depletions of cohesin have shown limited impact on chromatin organisation[19–21]

To determine the role of cohesin in interphase chromatin, we deleted *NIPBL/SCC2*, the factor which is necessary for cohesin loading on chromatin[22]. Since the turnover of chromosome-bound cohesin is ~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 Figure 2). Ten days after tamoxifen

injection, *Nipbl* expression was dramatically reduced (Fig. 1b) and led to a displacement of cohesin from the chromatin fraction to the soluble nuclear fraction, indicating loss of cohesin from chromosomes (Fig. 1c). This strong depletion of chromatin-bound cohesin is also observed by calibrated ChIP-seq[25] for RAD21 and SMC3 both genome-wide (Extended Data Figure 3) and at WT cohesin peaks (Fig. 1d–e). A conservative estimate indicates at least a 4- to 6-fold decrease in chromatin-associated cohesin. No particular pathological signs compared to control animals (mock-injected *Nipbl^flox/flox*; *Ttr-cre/Esr1* animals or tamoxifen-injected *Nipbl^+/+*; *Ttr-*

Fig. 1 | **Overview of the experimental design.** (a) S 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 absence of *Nipbl,* cohesin (represented by a triangular ring) is not loaded on chromatin. (**b**) Expression of *Nipbl* and *Gapdh* (control) by RT-qPCR in control (n=4) and ΔNipbl (n=6 for *Nipbl*; 4 for *Gapdh*) hepatocytes. Mean normalized gene expression (using *Pgk1* as internal control, and WT expression level set as 1) is displayed as mean and s.d and compared using unpaired two-sided *t*-test (95% CI control=[0.7033–1.297]; mutant=[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, SMC1) from the chromatin-bound fraction. H2B and TOP2B distribution serves 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 data source. (**d**) Stacked heatmaps of calibrated ChIP-seq signal (for Rad21 and SMC3) at WT 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 20kb resolution of WT (left), TAM control (middle) and ΔNipbl cells (right). Top – *cis* compartment tracks. Middle – *cis* and *trans* contact maps of chr16–18. (**g**) An example of short-range Hi-C map at chr16:25–32Mb with compartment tracks.
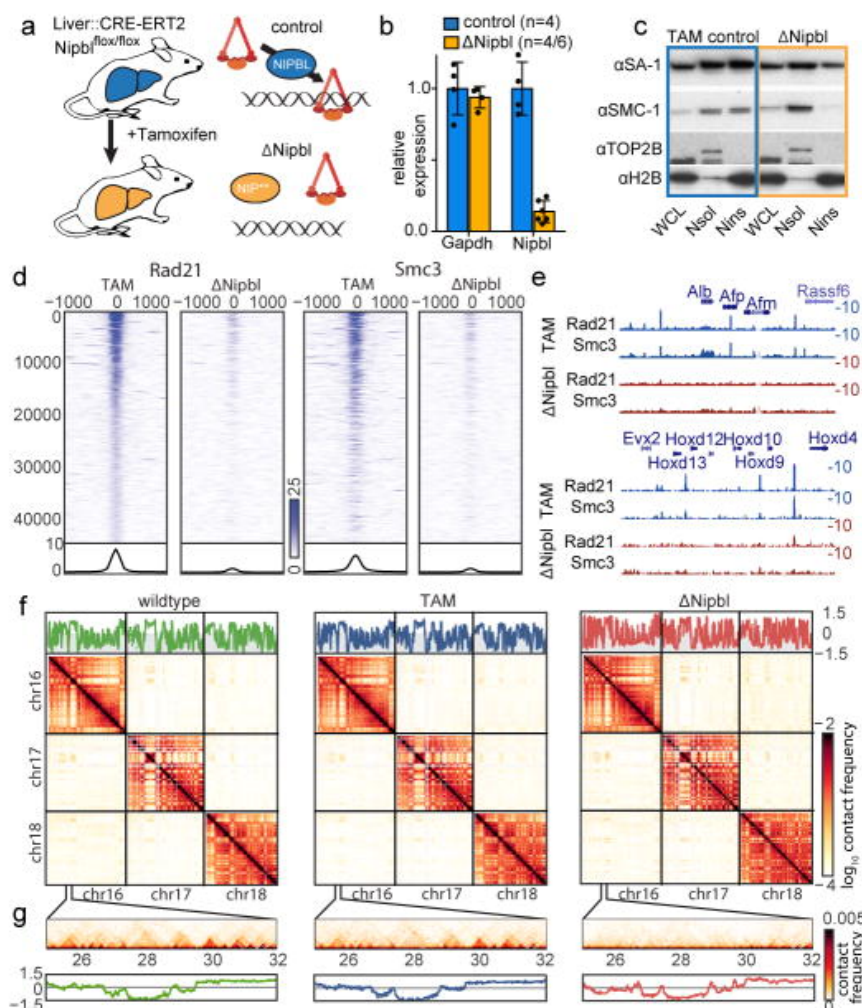
cre/Esr1 animals) were observed in the liver. Hepatocytes showed no sign of cell death or proliferation (Extended Data Figure 2).

To assess the consequence of *Nipbl* depletion and cohesin loss on chromosome organization, we performed tethered chromatin conformation capture[26] (referred below as Hi-C) on purified hepatocytes from wildtype (WT), tamoxifen control (TAM) and ΔNipbl animals (Fig. 1f). For each of these three conditions, two biological replicates were generated; 5 out of 6 replicates produced >25 million interactions at >10kb separation, on par with other primary tissue Hi-C ([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 datasets to generate Hi-C maps for the three different conditions. We compared Hi-C maps for ΔNipbl and control cells by examining compartments, TADs, peaks, and global

scaling of the contact probability $P(s)$[28] (Extended Data Figure 1).

Disappearance of TADs and peaks

Our Hi-C maps reveal a striking effect of *Nipbl* deletion on genome organization (Fig. 1f–g), contrasting with the very mild changes reported in previous cohesin depletion experiments (Extended Data Figure 4)[19–21]. Compared to WT and TAM control samples, ΔNipbl cells show genome-wide disappearance of local TAD patterns (Fig. 1g) but persistence of A/B compartmentalization (Fig. 1f). Disappearance of TADs in ΔNipbl is widespread and evident in individual maps (Fig. 2a), as well as on the composite map constructed by averaging the ΔNipbl Hi-C map around locations of domain boundaries detected in WT maps (Fig. 2c, Extended Data Figure 5). Some local
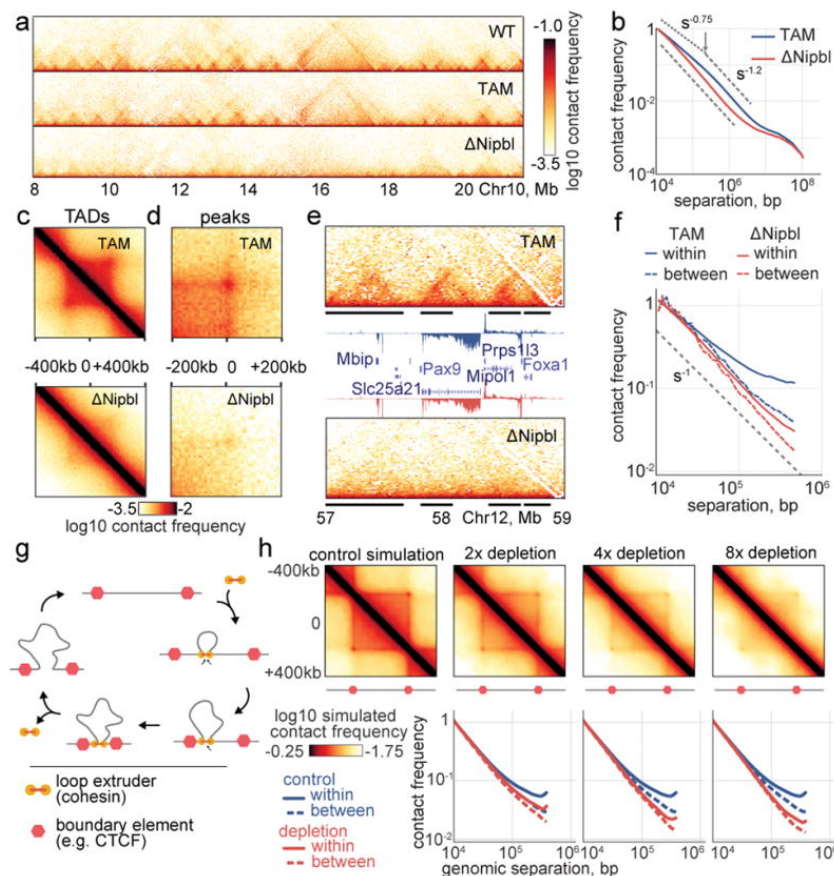

Fig. 2 | *Nipbl* deletion leads to disappearance of TADs and peaks from Hi-C contact maps. (a) The Hi-C map for chr10:8–21Mb illustrating loss of TADs and peaks. **(b)** Genome-wide P(s) curves in TAM control and ΔNipbl, normalized to unity at *s*=10kb. **(c, d)** The average Hi-C map of (c) 564 TADs of length 300–400kb and (d) 102 peaks of separation 500–600kb. **(e)** The Hi-C map of an example 2Mb region chr12:57–59Mb (top – TAM control, bottom – ΔNipbl) with expression tracks (middle – annotated genes and RPM normalized RNA-seq; sense above axis, antisense below; TAM in blue, ΔNipbl in red). Black bars show WT 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 300–500kb. **(g)** The 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 diagrams). After binding, cohesins translocate along the fiber in both directions, effectively extruding a loop (bottom diagrams). Extrusion halts when cohesins reach boundary elements. Extruded loops disassemble when cohesins unbind from the chromosome (left diagram). **(h)** Polymer simulations of loop extrusion reproduce the effects of cohesin depletion. Top row – average maps of TADs formed on contact maps in polymer simulations of loop extrusion. Left-to-right: the impact of sequential

organization is retained in regions with higher activity (A-compartment), where cohesin is about 3-fold more abundant than in inactive regions (B-compartment) in TAM (Extended Data Figures 5, 2e). We show below that these structures are not residual or new TADs, but unmasked small compartments. TADs-associated peaks of contact enrichment disappear as well in ΔNipbl maps, showing up to 4-fold reduction in contact frequency (Fig. 2d and Extended Data 4a), notably between convergent CTCF sites. Insulation and directionality of the contact footprint of CTCF sites also disappear upon *Nipbl* deletion (Extended Data Figure 6). However, we observe no effect of *Nipbl* deletion on CTCF occupancy, demonstrating that the loss of TADs and CTCF contact in ΔNipbl is not due to loss or changes in CTCF occupancy (Extended Data Figure 3), This further strengthens the very distinct roles of CTCF and cohesin in shaping chromosome architecture[14,18].

Importantly, these changes cannot be attributed to altered gene expression in ΔNipbl cells, as TAD patterns vanished equally in regions with unchanged expression (Fig. 2e) or with up- and down-regulated transcription (Extended Data Figure 5c). This major reorganization of chromatin architecture is also reflected in the curves of
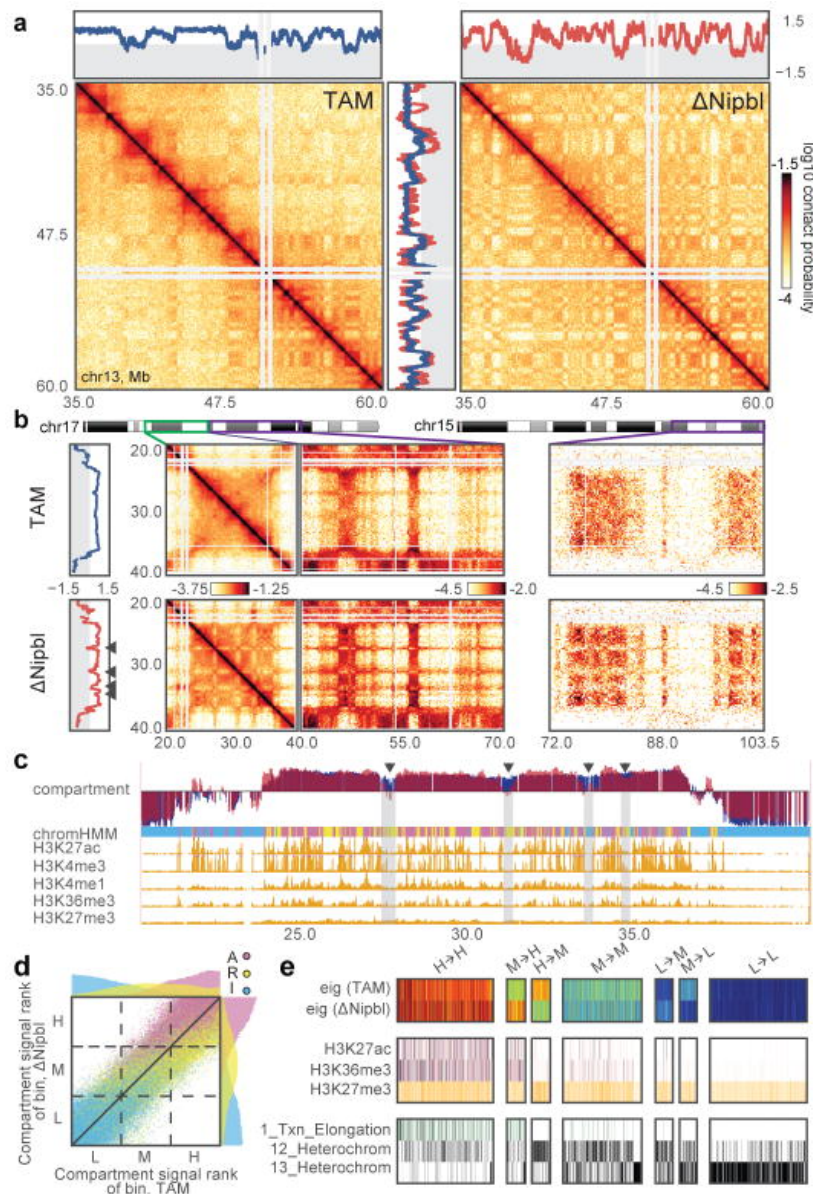

Fig. 3 | *Nipbl* deletion leads to activity-dependent alteration of compartment structure. (a) An example region (chr13:35–60Mb) showing changes in compartmentalization. Top and central panels – 20kb *cis* compartment tracks. **(b)** Hi-C maps of an example region at 200kb with predominantly positive (type A) compartment signal in TAM control (top) showing fragmentation upon *Nipbl* deletion (bottom), manifested in the alternating contact patterns of short- (<10Mb, middle left) and long-range *cis* (middle right) and *trans* contact maps (right panel). In (a–b), the two Hi-C replicates of each condition show similar results. **(c)** The loci experiencing a local drop are depleted in epigenetic marks of activity. Top to bottom: compartment track in TAM (blue) and ΔNipbl (red); simplified ChromHMM state assignments: active (magenta) / repressed (yellow)/inert (cyan); ENCODE activity-related histone ChIP-seq for adult mouse liver cells. In (b–c), arrowheads indicate local drops in compartment signal. **(d)** Rank correlation of 20kb compartment tracks (n=113,372 20kb genomic bins) in TAM and ΔNipbl, colored by simplified ChromHMM state. Top and right margins – histograms of compartment signal ranks split by simplified ChromHMM state in ΔNipbl (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 WT and ΔNipbl; 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, while bins transitioning from the high to the middle tercile were depleted in those marks.

contact frequency $P(s)$ as a function of genomic distance s [3,29]. In control samples, like in other mammalian cells[14,15], $P(s)$ curve has two regimes: a more shallow decay for $s<200$Kb ($P(s)\sim s^{-0.7}$), and a steeper scaling for 200Kb $< s < 3$Mb ($P(s)\sim s^{-1.2}$). Loss of chromatin-associated cohesin leads 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 WT TAD intervals: in controls, $P(s)$ within TADs decreases more slowly than $P(s)$ between TADs. In ΔNipbl, the within $P(s)$ curve collapses to the between $P(s)$ curve, indicating that the characteristic enrichment of contacts within TADs is lost and that chromatin folding becomes more uniform and decompacted (Fig. 2f), which is 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 simulations, 8-fold depletion is required to achieve agreement with our experimental data, manifested by (i) noticeable disappearance of TAD and corner peak enrichments; (ii) loss of the $P(s)\sim s^{-0.7}$ regime in the scaling; (iii) decompaction of chromatin (Fig. 2h). Together, these analyses and the observed effects of *Nipbl* deletion indicate that cohesin plays a central role in the local compaction of chromosomes, and support that this effect is mediated by the production of dynamic populations of extruded chromatin loops between boundary elements, which forms TAD and corner peak patterns in interphase Hi-C maps[14,15].

Enhanced and finer compartmentalization

We next examined compartmentalization of chromatin in ΔNipbl cells. As noted earlier, in contrast to the drastic loss of TADs, compartmentalization still exists (Fig. 1) and is enhanced (~1.8 fold) (Extended Data Figure 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 ΔNipbl cells with small B-like regions appearing inside A-regions (Fig. 3a,b, Extended Data Figure 7a). This finer compartmentalization is reflected in the shorter autocorrelation length of the compartment track (150Kb in ΔNipbl vs ~500Kb in WT and TAM (Extended Data Figure 7b). It explains most of the remaining or new domains and boundaries seen in ΔNipbl Hi-C maps (Extended Data Figures 8a). These emerging B-like regions possess the hallmarks of compartmentalization: (i) they are visible as local depressions in the ΔNipbl compartment track (Fig. 3b) and (ii) they show preferential interactions with other B-regions both in far *cis*- and in *trans*-chromosomal maps (Fig. 3b). Since 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. In contrast, predominantly B-type regions of the genome, do not show a similar fragmentation in ΔNipbl cells despite a complete loss of TAD patterns (Extended Data Figure 7a,b). Finally, Nipbl depletion reveals that WT TADs can span regions of opposing compartment type (Extended Data Figure 8b–d). Taken together, our observations defy the common notion of TADs simply being the building blocks of larger compartmental segments. Instead, TADs and compartments represent two independent, potentially antagonistic types of chromosomal organization.

Strikingly, we found that the compartmentalization profile of the ΔNipbl Hi-C map reflects local transcriptional activity and chromatin state better than that of the WT Hi-C map (Fig. 3c, Extended Data Figure 7c–d, 8c–d). The compartment track of ΔNipbl cells shows a stronger correlation with tracks of activity-associated epigenetic marks, e.g. H3K27ac, H3K4me3, DNase hypersensitivity, transcription factor binding etc., smoothed over a wide range of window sizes (Extended

Data Figure 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*). While 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 showing a reduction in compartment signal (from A to B-like), while those showing lamin-B1 association across different mouse cell lines are primarily B-type in both WT and the mutant (Extended Data Figure 9a, *Methods*). Importantly, these changes in compartmentalization cannot be attributed to changes in expression or in the activity marks (H3K27ac, H3K4me3) that are largely unperturbed in the mutant at the scales relevant to compartmentalization (Extended Data Figures 8d,9b–e). In summary, absence of cohesin enhances the compartmentalization of active and inactive chromatin: (i) A and B regions, as detected by Hi-C, form fewer contacts between one another, (ii) a finer compartment division emerges, (iii) this fine compartment structure corresponds better to the local functional states of the genome, even when considering those observed in WT. 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
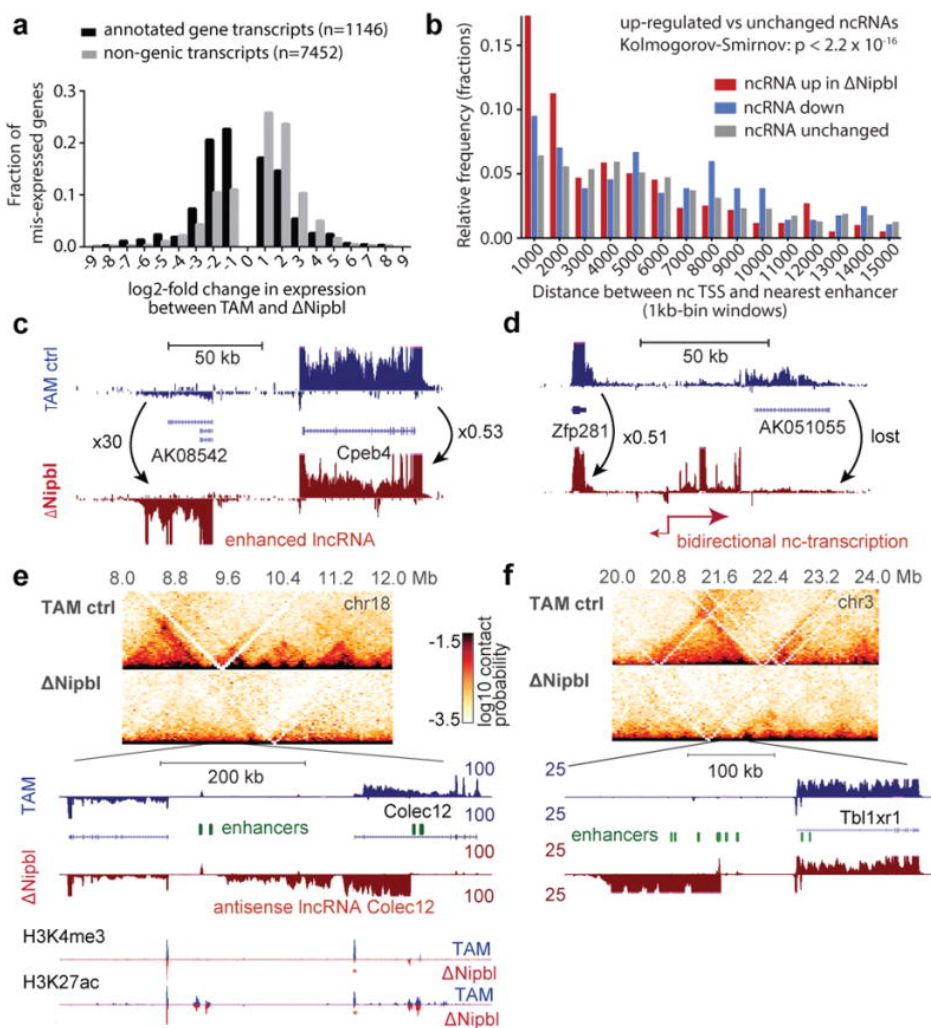

Fig. 4 | Transcriptional changes in *Nipbl* mutants reveal possible enhancer-promoter miscommunication in absence of TADs. (a) Distribution of fold-changes for genes (black, n=1146)) and exo-genic transcripts (light grey, n=7452) showing at least a two-fold change in expression. **(b–c)** Examples of transcriptional changes, with TAM control stranded RNA-seq tracks in blue, and ΔNipbl in red. Four replicates per condition, each confirming the reported changes, were combined. **(d)** Distribution of distances between the transcriptional start of the ncRNAs (up log2(foldchange)>3, n=595; down log2(foldchange)<−3, n=284; unchanged −0.5<log2(foldchange)<0.5, n=2238) and the nearest enhancer. **(e,f)** Switches in transcription at the *Colec12* and *Tbl1xr1* loci. Top panels show 40kb Hi-C maps of the 4Mb neighborhood. **(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), while peaks at the *Colec12* promoter disappear (red asterisk). **(f)** Exo-genic transcription emerges in ΔNipbl upstream of the *Tbl1xr1* gene on chr3.

observed a similar preservation of epigenetic state upon CTCF depletion[18].

Altogether, these results indicate that chromatin has an intrinsic tendency to form small-scale, specific compartments that reflect co-segregation based on the local epigenetic landscape and transcriptional activity, and that *Nipbl* and cohesin activity interfere with those clear subdivisions 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 are significantly mis-expressed (637 down-regulated, median fold-change=0.32, 487 up-regulated, median fold-change=3.15, with DESeq2 tools) upon *Nipbl* deletion (Fig. 4a and Extended Data Figure 10a–e). Gene ontology enrichment analysis does not give very strong indication of preferential effect on a biological function (Supplementary Table 4), reflecting possibly indirect and adaptive transcriptional changes.

While H3K27ac (and H3K4me3) peaks at promoters of affected genes change in coherence with expression changes, distal peaks (marking active distant enhancers) are mostly unaffected (Extended Data Figure 10f–i), indicating that while 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 down-regulated genes were surrounded by a larger intergenic space (defined by the distance separating the TSS of their immediate neighbors) than up-regulated or unaffected ones (Extended Data Figure 10b) and that transcriptional changes are concentrated within regions that form larger TADs in WT. This characteristic genomic context of transcriptional alterations is consistent with defective long-range regulatory interactions in ΔNipbl cells.

Closer examination of RNA-Seq tracks revealed a widespread up-regulation of exogenic (intergenic or antisense intragenic) transcription in ΔNipbl cells (Fig. 4a). Interestingly, while we found more genes down-regulated than up-regulated, the trend was opposite in exogenic regions. Using a conservative approach (see *Methods*), we found 1107 non-genic transcripts or transcribed regions, which showed at least an 8-fold enhanced transcription in ΔNipbl cells; among these, 232 corresponded to non-coding RNAs, which were not or barely detected in control samples, and often not annotated (Extended Data Figure 10d). The new transcription is often bi-directional, (Fig. 4c–d, Extended Data Figure 10f–j) and occurs either at small pre-exiting H3K4me3 peaks, corresponding likely to poised promoters, or at active H3K27ac enhancers (Fig. 4b, e, f, Extended Data Figure 10f–j). We saw several examples of reciprocal expression changes (i.e. down-regulation of a gene being followed by up-regulation of an adjacent gene or of a new non-coding transcripts) (Fig. 4c–d, Extended Data Figure 10f–j), but often, new non-coding transcription arises 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* loss impairs enhancer communication: with a reduced range of contact due to absence of TADs, some enhancers may not reach their target promoters and instead transfer their activity on nearby alternative, sometimes cryptic, targets (including themselves).

Cohesin is central to chromosome folding

Overall, our findings provide insights into the mechanisms that organize the interphase genome in 3D and their relation to gene expression. Our data shows the essential role of cohesin in the formation of TADs. It is possible that the cohesin depletion in previous studies that did not report such drastic effects[19–21] was insufficient to achieve substantial loss of TADs. Our simulations suggest that TADs would still be pronounced at 2-fold cohesin

depletion, requiring ~8-fold depletion for loss of TADs, which is close to what we observed in ΔNipbl samples. It is however also possible that *NIPBL* may affect cohesin activity at several levels, i.e. not only as a loader but also by facilitating ATP hydrolysis and loop extrusion[33], which could further impact TAD formation. This could also account for why a recent deletion of *Scc4/Mau2*, 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/Scc1) knockout zygotes demonstrated complete loss of TADs and associated peaks[35]. Emergence of fine compartmentalization, however, was not reported likely due to the limited number of nuclei. Two other preprints reporting degron-mediated depletion of Rad21/Scc1 in human cell lines observed concordant effects on TADs, peaks, and compartmentalization to our study[36,37]. Rao et

Fig. 5 | Two independent but overlapping modes of chromosomal 3D organization. TADs (colored 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 mutant Hi-C map relative to that of the WT and whose alternating member regions more faithfully track transcriptional activity. The resulting reduction of contact range (right) thwarts distant enhancers (ovals) from acting on their normal target genes (arrows, with colored ones indicating active genes, white ones inactive), leading them to act instead on neighboring genes or cryptic promoters located in their vicinity. The active units make up new compartmental regions (grey triangle).

al. also identified intriguing emergent "loop" cliques enriched in super-enhancers upon cohesin depletion[37]. We also observed such features in our primary hepatocytes using a new Hi-C browser[38], though we note that in both cases these patterns do not resemble the sharp *cis*-only focal patterns of CTCF-associated Hi-C peaks commonly referred to as loops[12]. Rather, these larger patches appear to be enhanced compartmental interactions between very active regions that line up well with actively 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. Remarkably, the changes we observed demonstrate that there are at least two mechanisms of independent origin whose overlapping action organizes mammalian interphase chromatin (Fig. 5). First, the global spatial compartmentalization of active and inactive regions of the genome is achieved by a cohesin-independent mechanism, which acts globally and across scales, including scales smaller than previously appreciated. This compartmentalization is however blurred by the action of a second, cohesin-dependent, mechanism that compacts chromatin locally, independently of its status. Noteworthy, the independence of the two mechanisms is supported the absence of compartments, despite of the existence of TADs, in maternal zygotic pronuclei [39].

The co-existence of two processes with different modes and scales of action can help 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[7,8,12,40–42]. Our data demonstrates clearly the existence of local, cohesin-dependent, self-interacting domains identifiable as TADs. The experimental ability to now distinguish these two modes of chromosome organization provides avenues to
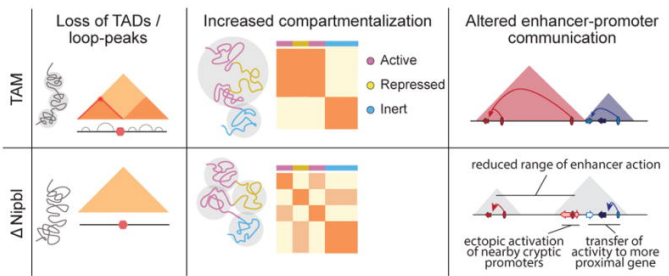
dissect the process(es) governing their formation and maintenance, as well as to characterize their relationship to gene expression. In this respect, the ability of cohesin to bring regions of different activities together may play an essential role in initiating changes in gene expression driven by distant enhancers, while compartmentalization may maintain existing regulatory interactions.

Experimental procedures

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 cells. Individual ESC clones were screened for successful recombination by Southern blotting deploying two unique probes hybridizing 5′ and 3′ to the integration site, respectively. Cells of a successful clone were injected into mouse blastocysts and resultant chimera were bred to C57BL/6J mice. Offspring were analyzed for successful germ line transmission by PCR and Southern blotting. *Nipbl^flox/+* mice were maintained on C57BL/6J genetic background.

For acute deletion of the *floxed* exon, we used either a constitutive ubiquitous CRE-driver (Hprt::Cre[43]), a limb-specific CRE-driver (Prx1::Cre[44]) or an inducible, liver-specific *Cre* allele (*Ttr-cre/Esr1*[45]).

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 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* [45]. Resultant double heterozygous mice were back-crossed to *Nipbl^flox/flox*. For experiments we used animals homozygous for the floxed *Nipbl* allele either carrying 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*), respectively.

12 week old mice were injected with 1mg Tamoxifen (100μl of 10mg/ml Tamoxifen in corn oil) on 5 consecutive days. After keeping these mice for another 5 days without injection, they were sacrificed and the hepatocytes were harvested. Until this time point, mice displayed no abnormal behavior, weight loss or any other obvious physiological changes. This was also the case, when we kept mice for additional 4 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 prepped for a two-step perfusion adapted from[46,47]. First, the liver is perfused with an EDTA-containing buffer removing 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 (60g for 1 min), to reduce non-mesenchymal cells and debris, hence, enriching intact hepatocytes in the sample. Part of each sample was fixed with 1% PFA for 10 minutes at room temperature. Fixed and unfixed hepatocytes were aliquoted and frozen in liN$_2$ 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 following primers: Nipbl-qPCR_F TCCCCAGTATGACCCTGTTT, Nipbl-qPCR_R AGAACATTTAGCCCGTTTGG, Gapdh-qPCR_F CTCCCACTCTTCCACCTTCG, Gapdh-qPCR_R CCACCACCCTGTTGCTGTAG, RTqPCR_Pgk1_Fwd TGGTATACCTGCTGGCTGGATGG and RTqPCR_Pgk1_Rev GACCCACAGCCTCGGCATATTTC.

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 (a courtesy of Ana Losada, CNIO, Madrid) and Topo IIβ (611492, BD Biosciences) and Histone H2B (07-371, Millipore) as control for nuclear soluble and nuclear insoluble fractions, respectively.

Immunohistochemistry on liver

Slices of adult livers were collected and fixed in 4% PFA overnight. After dehydration, the tissues were embedded in paraffin and sectioned at 6μm. The sections were deparaffinized with xylene, rehydrated and antigens were retrieved by boiling in citrate buffer. The sections were blocked in 10% FBS and incubated with primary antibodies (αphospho-Histone H3, 06-570 Millipore; αcleaved-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) prior to 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) their size-distributions were determined with Bioanalyzer. Equimolar pools of libraries were sequenced with Illumina HiSeq2000 (50bp, single end). We retrieved on average 25 mio reads per sample, of which 19 mio 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. 20μg chromatin per sample was diluted in RIPA and incubated with 1.5μg of either αH3k4me3 antibody (C15410003-50, Diagenode) or αH3K27Ac 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 ChIPped 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) their size-distributions were determined with Bioanalyzer. Equimolar pools of libraries were sequenced with Illumina HiSeq2000 (50bp, single end). We retrieved on average 20 mio reads per sample, of which 15 mio 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 according to ref. [52]. Briefly, 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 minutes prior to transfer to shearing cuvette (Covaris-520130). Lysis was performed in a Covaris E220 sonicator for 5 minutes according to the following parameters: Peak power = 75W, Duty factor = 2%, 200 cycles/burst, temperature = 4C. Following lysis, nuclei were pelleted in a chilled 4C centrifuge at 1000g for 5 minutes 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 minutes with the following settings: Peak power = 140W, Duty factor = 5%, 200 cycles/burst, temperature = 4C. 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 100μl of Dynabeads M-280 sheep anti-rabbit. Chromatin Immunoprecipitations were performed successively (Rad21, followed by CTCF, followed by SMC3) using unbound chromatin from the first IP to perform the next one. As IP efficiency is very low for this protein, we did not observe a change of signal when comparing sequential to parallel IP performed on control chromatin. Antibody-bound beads were washed 2x in ice cold PBS and incubated with 75ug of sheared mouse hepatocyte chromatin pooled with 75ug HEK293 human chromatin (as internal control and calibration) in RIPA buffer (50mM TrisHCl, pH 8.0, 0.15M NaCl, 1% Triton X-100 and 0.1% sodium deoxycholate) overnight at 4C on a rotating wheel and then subsequently washed 2x 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 50ug/mL RNaseA and incubated for 30 min at 37C, followed by reverse cross-linking with 0.5 mg/ml Proteinase K in TE supplemented with 0.5% SDS overnight at 65C in a shaking thermomixer at 1200rpm. 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 which chromatin is used as internal control for the IP were not authenticated, nor screened for mycoplasma contamination.

Tethered Chromatin Capture (TCC)

Roughly 100 mio fixed hepatocytes per sample were processed according to Kalhor et al.[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 (50bp, paired end). We retrieved between 100 and 150 mio 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 in[48] and implemented in the *hiclib* library for Python (publicly available at https://bitbucket.org/mirnylab/hiclib). Upon filtering PCR duplicates and reads mapped to multiple or zero locations, we aggregated the reads pairs into 20kb and 100kb 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 each genomic bin follows a lognormal distribution. This shape of a distribution is expected when different factors (GC content, mappability, restriction fragment density, etc) affect the visibility of a genomic bin in a multiplicative fashion. The MAD-max procedure has three steps:

1. Normalize the numbers of contacts at each genomic bin via dividing them by their corresponding chromosome-wide medians. This step accounts for a possible variation in chromosome copy numbers, e.g. in chrX.

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 center of the distribution as the median of log(normalized contact numbers), and the width as the Median Absolute Deviation from this median, or, MAD.

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

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

To remove the short-range Hi-C artifacts - unligated and self-ligated Hi-C molecules - we removed contacts mapping to the same or adjacent genomic bins. The filtered 20kb and 100kb contacts matrices were then normalized using the iterative correction procedure (IC)

[48], such that the genome-wide sum of contact probability for each row/column equals 1.0. Observed/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 analyze other existing Hi-C datasets on cohesin-depleted cells[19–21].

Compartment analysis via eigenvector decomposition

The compartment structure of Hi-C maps was detected using a modified procedure from[48]. In short, compartment tracks (i.e. the values of compartment signal across all genomic bins) were quantified as the dominant eigenvector of the observed/expected 20kb and 100kb *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 $\sqrt{AA*BB/AB2}$, 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 [49], 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 from [12] and the other was based on network modularity [50], 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 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[51]. 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/nezar-compbio/lavaburst).

The domain calls used in composite heatmaps were computed using the corner score on 20kb 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 20kb WT and ΔNipbl contact matrices sweeping over a range of gamma values to obtain a 1D boundary (i.e., 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 CH12-LX mouse cell line in[12].

Analysis of epigenetic features

To study the influence of the epigenetic chromatin state on genomic architecture, we used the ChromHMM genomic state annotation from[52]. Briefly, 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 loci one of 15 possible epigenetic states. For our analyses, we further grouped these 15 states into three: Active (states 1–11, 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[53] in the mm9 genome (filtered by p-value > 0.0001), intersected with CTCF binding peaks in ENCODE adult mouse liver dataset[54]. This procedure yielded 27840 peaks.

To study how lamina association affects genome compartmentalization, we used the dataset of Lamin-B1-binding loci from[55] containing data from four mouse cell lines: embryonic stem cells (ESC), multipotent neural precursor cells (NPC), terminally differentiated astrocytes (AC) and embryonic fibroblasts (MEF). We then selected 100kb 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 for 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 mm9 reference mouse genome assembly and GENCODE vM1 transcriptome[56] using STAR v.2.5.0a[57] 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[58] with the "union" option and called DE genes with DESeq2[59].

For identification of exogenic transcripts, we aligned reads on the mm9 genome using HISAT2[60] with the option to authorize splicing events. From these alignments, we count the number of reads in sliding windows of 1kb (with a step of 600nt) taking each strand of the genome as separate. We use a gaussian mixture model (2 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 7 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 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 datasets. 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 ΔNipbl conditions to "calibrate" the raw mm9 signal pileups in order to make the binding profiles quantitatively comparable. Briefly, for each factor (CTCF, Rad21, Smc3), we computed genome-wide binned coverage at 200bp 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 Figure 6b), we used the ratio of the top 0.2% of 200bp bins in TAM vs ΔNipbl to rescale the mm9 signal in the ΔNipbl 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 datasets using alignments of mm9-mapped input DNA as background.

GO enrichment analysis

GO enrichment analysis was performed on significantly differentially expressed genes (DESeq2 adjusted p-value < 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 ΔNipbl condition. ED Table 1 displays the list of enriched GO terms with a Benjamini-Hochberg p-value < 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 previously [14]. We then generated contact maps and calculated P(s), also as previously described.

We considered a system of 42 consecutive TADs of 400kb, where impermeable BEs were placed between each pair of neighboring TADs. We modeled the chromatin fiber as a series of 1kb monomers (~6 nucleosomes, ~20nm), 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[14,61].

Three-dimensional polymer dynamics were implemented using OpenMM, a high-performance GPU-assisted molecular dynamics API[62,63]. Simulations were initialized from a compact polymer conformation, as described in[64], 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 fiber to equilibrate with this set of LEF-imposed bonds, simulations were then advanced for 400 blocks of simulations. After that, 2000 blocks (ie. 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 5000 conformations for each parameter set were obtained.

For the control simulation, we considered LEFs with 200kb processivity and 400kb separation. To investigate the effect of depleting the amount of bound cohesin, we then increased LEF separation by 2-fold, 4-fold, and 8-fold. All simulations used a stiffness of 2, density of 0.2, 3D-to-1D-steps of 2500. 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 (10kb) bins. Then, the map was normalized such the average value of the first diagonal equals one, and log10 transformed. Finally, the color-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 10kb was equal to 1.

Supplementary Information

Two independent modes of chromatin organization revealed by cohesin removal

Wibke Schwarzer, Nezar Abdennur, Anton Goloborodko, et al.

Supplementary Tables.

Supplementary Table 1 – Summary statistics of read pairs for TCC (Hi-C) libraries in this study.

Supplementary Table 2 – Statistics of Hi-C libraries in Schmitt et.al, 2016[27], an example of a contemporary study on primary tissues.

Supplementary Table 3 – Statistics of Hi-C libraries in other cohesin depletion studies.

Supplementary Table 4 - The list of enriched GO terms for differentially expressed genes, with a Benjamini-Hochberg p-value < 0.05.

Supplementary Table 5 – Public data sources

Supplementary Data 1.

Original scans of western blots. The selected regions for the different panels shown on Fig.1b are boxed in red. Additional biological replicates of the two conditions can be seen (samples 1 and 4)
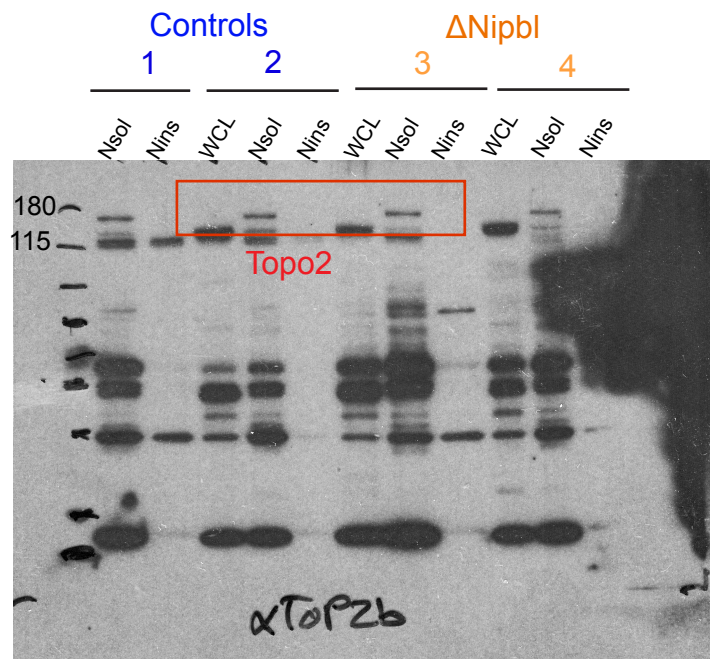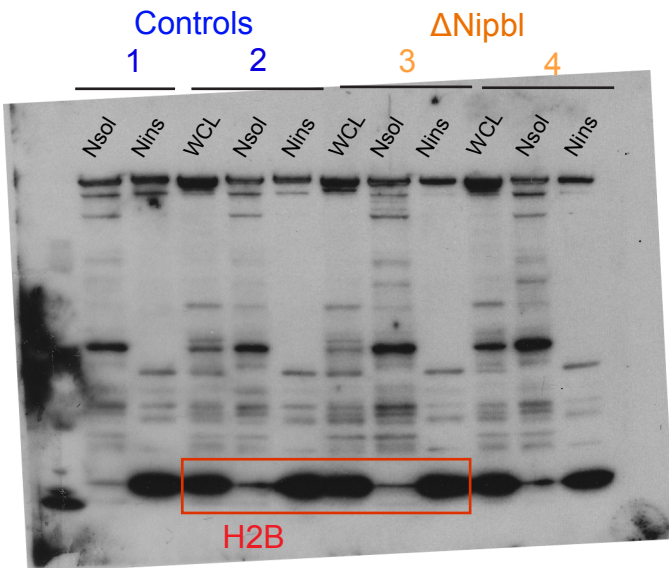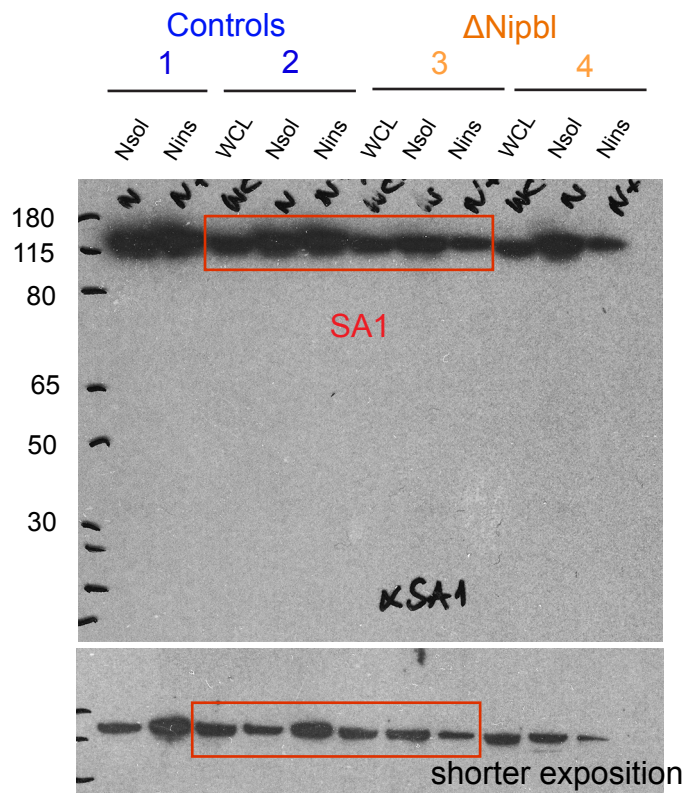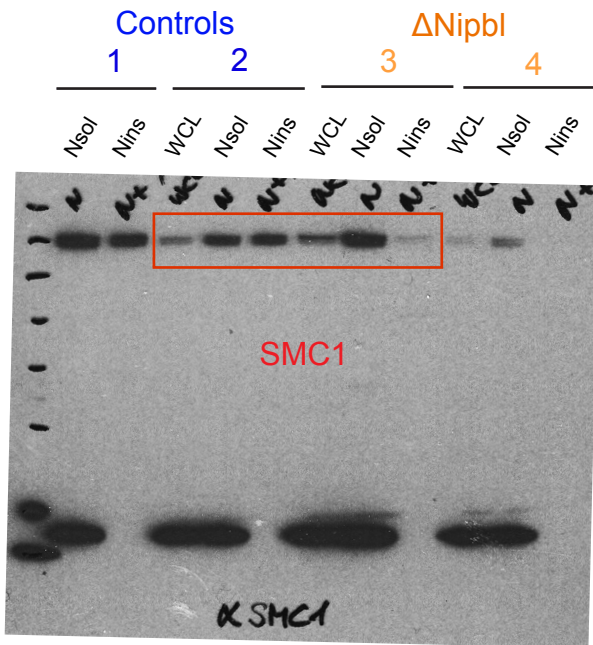
Supplementary Data 2.

Experimental Hi-C maps show extensive similarities between replicates. (a) The curves of contact frequency P(s) vs genomic distance s. (b) Pearson correlation coefficients between *cis* eigenvectors at 100kb. (c) The average Hi-C maps around 102 500kb-600kb peaks. (d) Compartmentalization saddle plots.

Supplementary Data 3.

Consistent differences in compartment eigenvector upon Nipbl deletion (a) The genome-wide *cis* eigenvectors at 100kb: WT (green), TAM (blue), ΔNipbl (red) are shown. (b) Zoom-in of the region shaded in grey in (a) using 20kb eigenvectors. In (a-b), the curves are shown for each of two replicates of each condition.
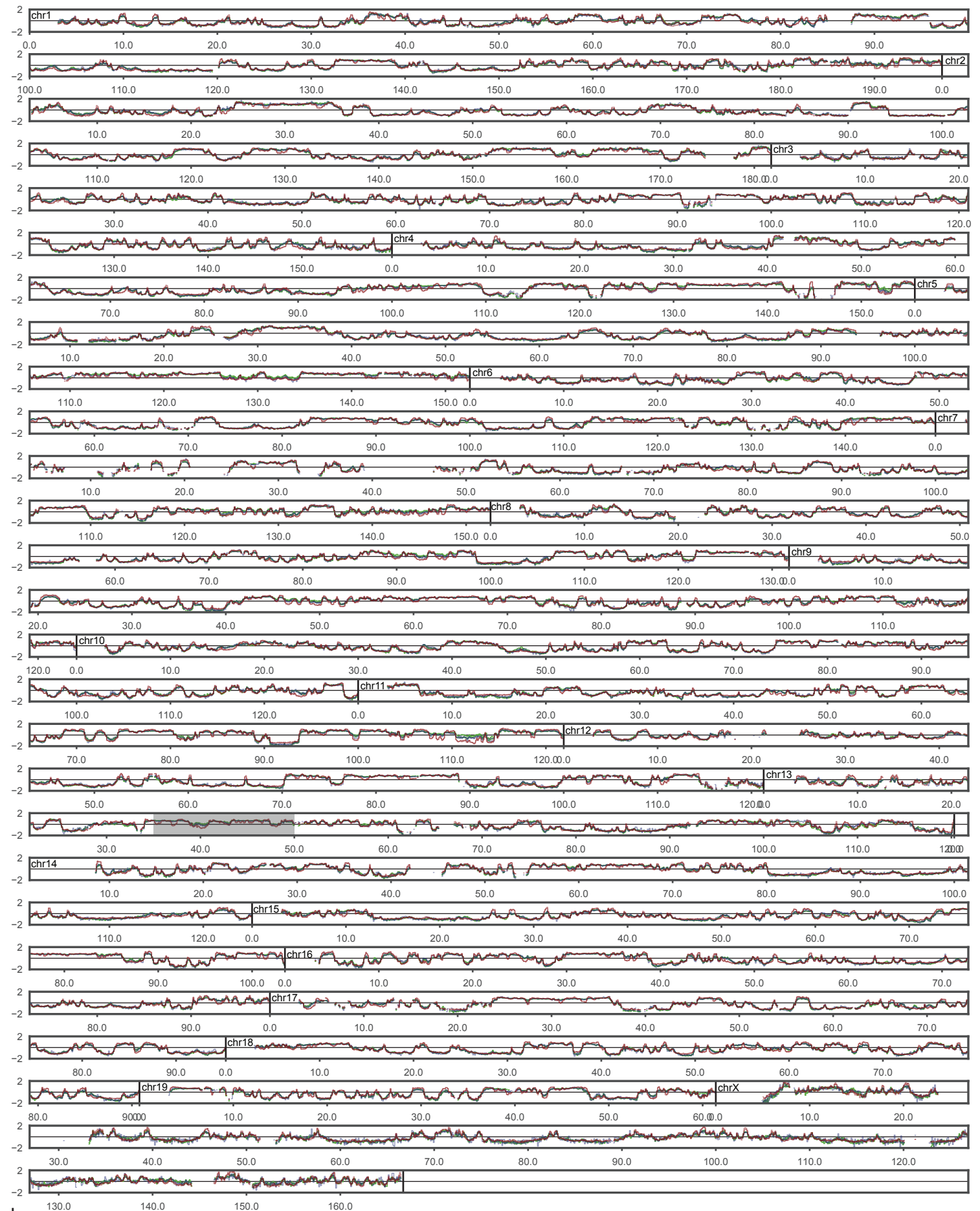
Supplementary Data 1

Supplementary Data 2
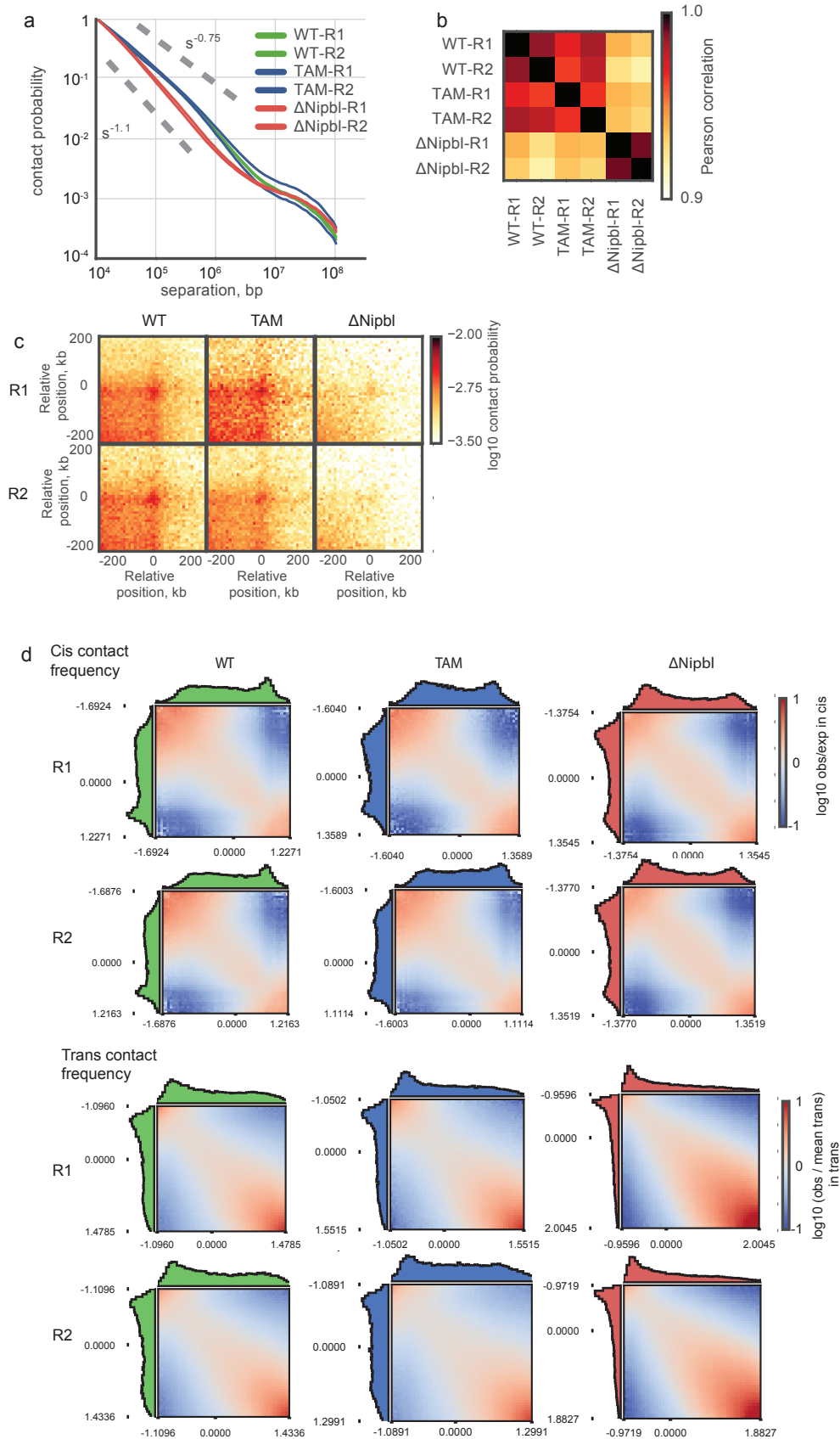
Supplementary Data 3

a

b
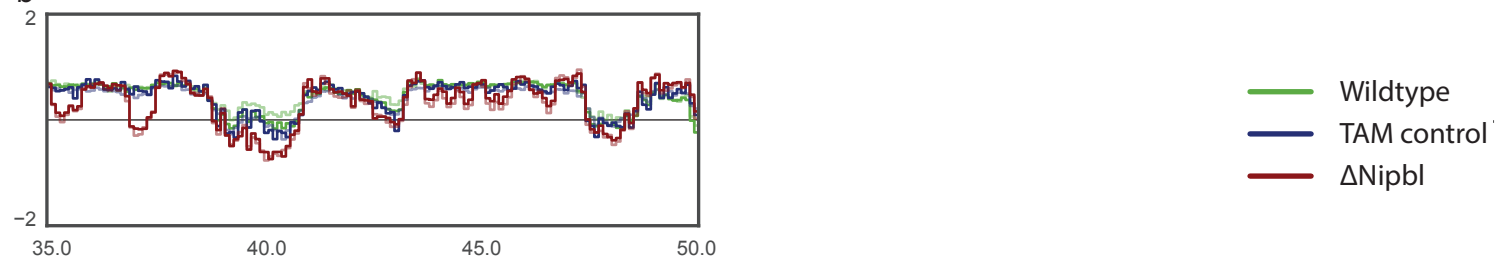
Wildtype

TAM control

ΔNipbl

Supplementary Table 1 – Summary statistics of read pairs for TCC (Hi-C) libraries in this study.

Replicate	codename	Raw read pairs	Mapped read pairs	Filtered read pairs	Cis read pairs, separation >10kb
WT-R1	014637Untr	139,343,707	75,465,383	51,745,705	35,529,637
WT-R2	019440Untr	114,258,179	60,638,254	43,097,410	28,704,961
TAM-R1	014199TAMcontr	94,032,309	53,285,208	14,530,511	9,870,917
TAM-R2	014399TAMcontr	134,045,253	72,789,581	51,281,408	34,490,881
ΔNipbl-R1	014200Nipbl	131,409,162	66,635,374	42,405,784	26,633,610
ΔNipbl-R2	016350Nipbl	133,914,046	69,894,091	50,527,104	31,561,513

Supplementary Table 2 – Statistics of Hi-C libraries in Schmitt et.al, 2016, an example of a contemporary study on primary tissues.

Study	Raw read pairs	Filtered pairs	Cis read pairs, separation >10kb
Bladder, R1	233,041,838	52,456,550	24,121,444
Bladder, R2	236,250,557	43,395,294	19,659,511
Adrenal Gland, R1	181,680,303	48,951,415	22,449,886
Psoas Muscle, R1	196,863,028	36,856,215	13,566,008
Psoas Muscle, R2	95,110,351	25,776,091	10,410,829
Pancreas, R1	160,416,883	49,627,891	21,200,091
Pancreas, R2	12,147,138	3,538,979	1,367,214
Pancreas, R3	128,745,528	15,683,831	7,787,867
Lung, R1	308,335,703	85,597,004	24,417,652
Hippocampus, R1	187,658,680	64,151,770	29,220,961
Right Ventricle, R1	672,319,838	179,219,920	63,091,486
Small Bowel, R1	177,864,628	58,017,249	21,073,296
Spleen, R1	378,529,216	107,599,392	33,032,593
Ovary, R1	201,215,217	70,283,904	20,909,766
Dorsolateral Prefrontal Cortex, R1	165,728,687	52,713,868	25,426,433

Supplementary Table 3 – Statistics of Hi-C libraries in other cohesin depletion studies.

Study	Filtered read pairs	Cis read pairs, separation >10kb
Seitan et al, 2013, WT, R1	51,337,902	12,676,930
Seitan et al, 2013, WT, R2	96,784,574	28,679,223
Seitan et al, 2013, ΔRad21, R1	57,060,666	14,219,809
Seitan et al, 2013, ΔRad21, R2	96,784,574	22,594,749
Zuin et.al, 2014, Control, R1	120,366,592	48,221,154
Zuin et.al, 2014, Control, R2	134,277,024	46,378,835
Zuin et.al, 2014, Rad 21 cleaved, R1	113,394,855	41,623,810
Zuin et.al, 2014, Rad 21 cleaved, R2	159,090,030	51,870,950
Sofueva et.al, 2012, Control NSC, R1	16,886,965	12,761,787
Sofueva et.al, 2012, ΔRad21 NSC Lox, R1	69,788,779	52,087,400
Sofueva et.al, 2012, ΔRad21 NSC Lox, R2	63,347,539	47,760,318
Sofueva et.al, 2012, Control AST, R1	29,246,903	21,885,386
Sofueva et.al, 2012, ΔRad21 AST Lox, R1	24,710,993	15,866,342
Sofueva et.al, 2012, ΔRad21 AST Lox, R2	20,287,207	13,980,244

Supplementary Table 4 – The list of enriched GO terms for differentially expressed genes, with a Benjamini-Hochberg p-value < 0.05.

Term	Category	Name	Regulated	Count	%
GO:0060271	BP	cilium morphogenesis	up	16	3.6
GO:0070062	CC	extracellular exosome	up	107	23.8
GO:0042995	CC	cell projection	up	36	8.0
GO:0005929	CC	cilium	up	18	4.0
GO:0005930	CC	axoneme	up	10	2.2
GO:0032869	BP	cellular response to insulin stimulus	down	14	2.4
GO:0006629	BP	lipid metabolic process	down	35	6.0
GO:0005615	CC	extracellular space	down	83	14.0
GO:0016021	CC	integral component of membrane	down	231	39.1
GO:0005783	CC	endoplasmic reticulum	down	88	14.9
GO:0005576	CC	extracellular region	down	84	14.2
GO:0005789	CC	endoplasmic reticulum membrane	down	54	9.1
GO:0043231	CC	intracellular membrane-bounded organelle	down	48	8.1
GO:0016020	CC	membrane	down	273	46.2
GO:0016324	CC	apical plasma membrane	down	25	4.2
GO:0005887	CC	integral component of plasma membrane	down	51	8.6
GO:0020037	MF	heme binding	down	21	3.6
GO:0016491	MF	oxidoreductase activity	down	47	8.0
GO:0052689	MF	carboxylic ester hydrolase activity	down	11	1.9

Supplementary Table 5 – Public data sources

Data type	Description	Cell-type	Ref.	URL
Hi-C data	Hi-C raw reads	Mouse ESC, NSC, AST	18	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE49018
	Hi-C raw reads	Mouse Developing Thymocytes	19	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48763
	Hi-C raw reads	Human HEK293T	17	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE44267
Peaks	Annotations from Hi-C	Mouse CH12-LX	8	ftp://ftp.ncbi.nlm.nih.gov/geo/series/GSE63nnn/GSE63525/suppl/GSE63525_CH12-LX_HiCCUPS_looplist.txt.gz
CTCF M1 motif	PWM for mouse		56	ftp://ftp.ebi.ac.uk/pub/databases/vertebrategenomics/FOG03/calls/CTCF_canonical_species
Lamin-B1 DamID	Normalized signal Normalized signal Normalized signal Normalized signal	Mouse ESCs NPCs Astrocytes MEFs	37	http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/laminB1_AC.txt.gz http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/laminB1_EF.txt.gz http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/laminB1_ES.txt.gz http://hgdownload.cse.ucsc.edu/goldenPath/mm9/database/laminB1_NP.txt.gz
	HMM state calls		38	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE17051
ChromHMM	Chromatin state segmentation trained on ENCODE mouse ChIP-seq data	Mouse Liver, Adult8wks	55	https://github.com/gireeshkbogu/chromatin_states_chromHMM_mm9
ENCODE Histone ChIP-seq	ChIP-seq, peak and bigwig files Input H3K27ac H3K27me3 H3K4me3 H3K4me1 H3K9me3 H3K36me3 H3K79me2	Mouse Liver, Adult8wks	45 and ENCODE	http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeLicrHistone/
ENCODE TFBS ChIP-seq	ChIP-seq, Peak and bigwig files Input PolII CTCF	Mouse Liver, Adult8wks	45 and ENCODE	http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeLicrTfbs/
ENCODE DNase-seq	DNAse I HS, bigwig	Mouse Liver, Adult8wks	45 and ENCODE	http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeUwDnase/
ENCODE RNA-seq	Poly-A RNA-seq, bigwig Poly-A RNA-seq, bigwig	Mouse Liver, Adult8wks Mouse Liver, Adult8wks	45 and ENCODE	http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeUwRnaSeq/ http://hgdownload.cse.ucsc.edu/goldenPath/mm9/encodeDCC/wgEncodeCshlLongRnaSeq/