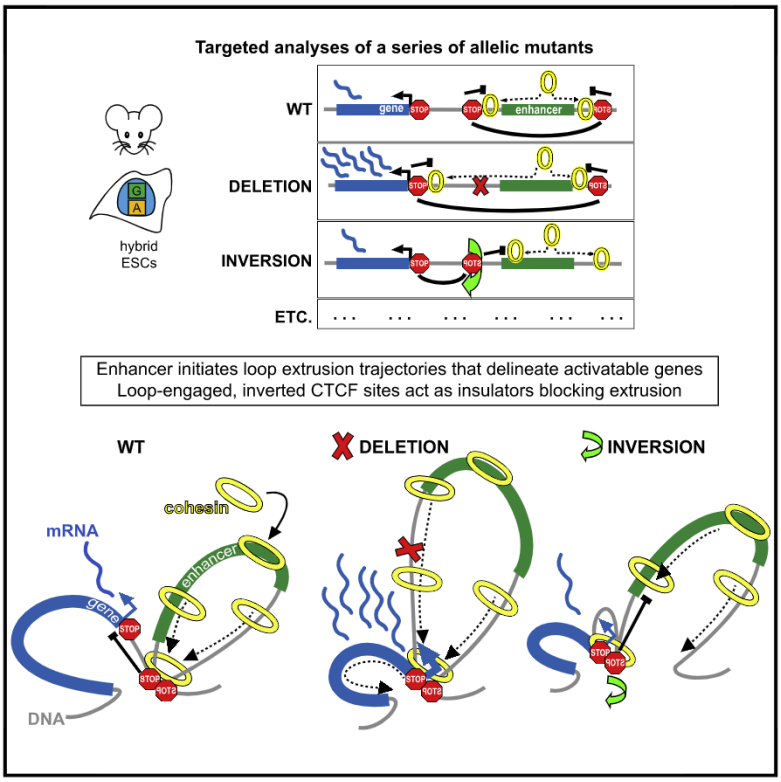
Molecular Cell

Article

Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression

Graphical abstract



Authors

Erica S.M. Vos,
Christian Valdes-Quezada,
Yike Huang, ..., Esther C.H. Uijttewaal,
Peter H.L. Krijger, Wouter de Laat

Correspondence

w.delaat@hubrecht.eu

In brief

The three-dimensional structure of the genome, organized by architectural proteins CTCF and cohesin, has a strong influence on how genes are expressed. Here, Vos et al. allele-specifically dissect an active genomic region to reveal how an enhancer and CTCF boundaries organize cohesin extrusion trajectories, build structural domains, and regulate genes.

Highlights

- An enhancer initiates loop extrusion trajectories that delineate activatable genes
- Enhancer insulation does not depend strictly on looping between its CTCF boundaries
- A loop-engaged inverted CTCF boundary can also block cohesin extrusion trajectories
- A strong CTCF boundary permits rare but productive enhancer-promoter contacts

Vos et al., 2021, Molecular Cell *81*, 3082–3095
August 5, 2021
https://doi.org/10.1016/j.molcel.2021.06.008

CellPress

CellPress

Molecular Cell

Article

Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression

Erica S.M. Vos,[1] Christian Valdes-Quezada,[1,2] Yike Huang,[1,2] Amin Allahyar,[1] Marjon J.A.M. Verstegen,[1] Anna-Karina Felder,[1] Floor van der Vegt,[1] Esther C.H. Uijttewaal,[1] Peter H.L. Krijger,[1] and Wouter de Laat[1,3,*]

[1]Oncode Institute, Hubrecht Institute-KNAW and University Medical Center Utrecht, 3584 CT Utrecht, the Netherlands

[2]These authors contributed equally

[3]Lead contact

*Correspondence: w.delaat@hubrecht.eu

https://doi.org/10.1016/j.molcel.2021.06.008

SUMMARY

To understand how chromatin domains coordinate gene expression, we dissected select genetic elements organizing topology and transcription around the *Prdm14* super enhancer in mouse embryonic stem cells. Taking advantage of allelic polymorphisms, we developed methods to sensitively analyze changes in chromatin topology, gene expression, and protein recruitment. We show that enhancer insulation does not rely strictly on loop formation between its flanking boundaries, that the enhancer activates the *Slco5a1* gene beyond its prominent domain boundary, and that it recruits cohesin for loop extrusion. Upon boundary inversion, we find that oppositely oriented CTCF terminates extrusion trajectories but does not stall cohesin, while deleted or mutated CTCF sites allow cohesin to extend its trajectory. Enhancer-mediated gene activation occurs independent of paused loop extrusion near the gene promoter. We expand upon the loop extrusion model to propose that cohesin loading and extrusion trajectories originating at an enhancer contribute to gene activation.

INTRODUCTION

The spatial organization of the genome plays an important role in specifying transcription programs, yet how individual genetic elements such as enhancers, promoters, and insulators coordinate this process remains enigmatic. Enhancers tissue-specifically activate genes by physically approaching them, forming chromatin loops (Kagey et al., 2010; Splinter and de Laat, 2011; Tolhuis et al., 2002). Hereby it is thought that the local concentration of transcription factors, co-activators, and RNA polymerase II at promoters increases, thereby increasing the output of target genes. Active enhancers are acetylated at histone H3 lysine 27 (H3K27ac) (Creyghton et al., 2010) and vary in size from hundreds to many thousands of base pairs, while particularly long, highly active enhancers are termed super enhancers (SEs). These recruit large amounts of master regulator transcription factors and Mediator and are thereby thought to help define cell identity (Whyte et al., 2013). Ectopic enhancer-promoter interactions, such as those that occur in oncogenic translocations (Gröschel et al., 2014; Hnisz et al., 2016), are responsible for increased transcription of unintended target genes, in which enhancers are adopted, or "hijacked," by newly accessible gene promoters (Lettice et al., 2011; Northcott et al., 2014).

Looping between insulator elements seems to counteract the ability of enhancers to loop to non-target genes, thereby constraining enhancer activity. This type of looping between the end points, or boundaries, of chromatin domains forms topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012; Rao et al., 2014; Sexton et al., 2012), with enhancer-promoter looping generally found within these insulated structures (Dixon et al., 2016; Dowen et al., 2014; Sun et al., 2019; Weintraub et al., 2017). TAD boundaries are enriched for the ring-shaped molecule cohesin, which is thought to help form and/or stabilize looping by embracing interacting strands of DNA and by translocating along chromatin. As posited by the loop extrusion model (Fudenberg et al., 2017; Sanborn et al., 2015), cohesin's translocation away from its loading sites is halted by the versatile architectural protein, insulator, and transcription factor CCCTC-binding factor (CTCF), but only when this protein is properly oriented (Guo et al., 2015; Li et al., 2020; Rao et al., 2014; de Wit et al., 2015). CTCF mediates looping (Splinter et al., 2006) and is enriched at domain boundaries (Dixon et al., 2012), though it is also found within domains and sometimes even at gene promoters, within gene bodies, and at enhancers, thereby also apparently playing more direct roles in transcriptional regulation (Hanssen et al., 2017; Kubo et al., 2020; Nora et al., 2017; Ren et al., 2017; Ruiz-Velasco et al., 2017). Whereas CTCF binds specific DNA motifs directly, cohesin is deposited onto chromatin by the protein NIPBL (Ciosk et al., 2000; Newkirk et al., 2017), which is enriched at promoters and (super) enhancers (Dowen et al., 2013; Hnisz et al., 2013; Kagey et al., 2010). Therefore SEs, because of their multiple NIPBL sites,

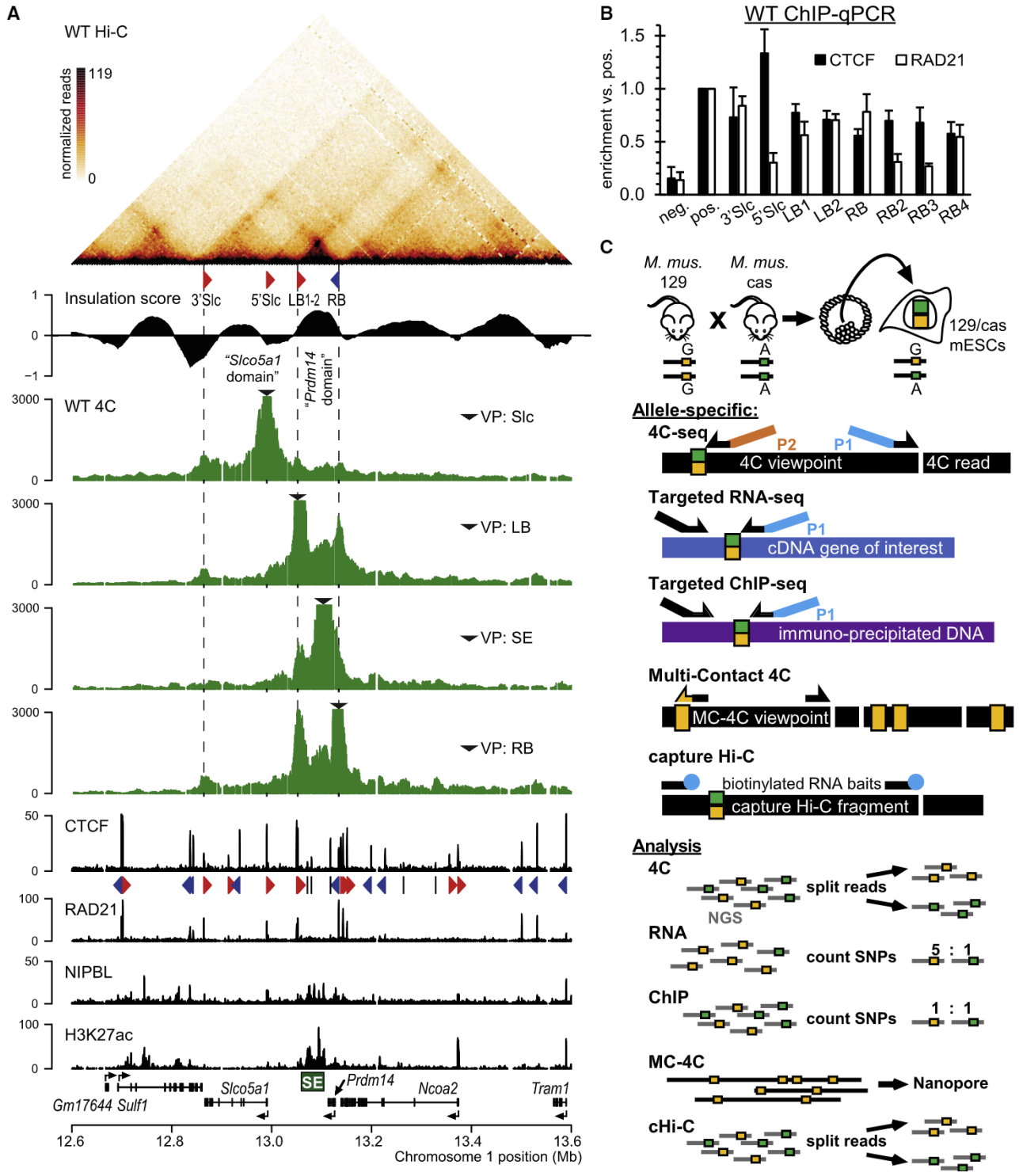


Figure 1. Chromatin features of the *Prdm14* locus in wild-type mESCs

(A) Top: Hi-C heatmap (Bonev et al., 2017); KR normalized reads. Note high interaction frequency between CTCF sites LB1-2 and RB. Below Hi-C, insulation score across the region. Middle: WT 4C profiles reveal detailed interactions across the locus for each viewpoint (VP). y axis, 4C coverage per 1 million normalized reads. Bottom: available ChIP-seq tracks for CTCF (Nora et al., 2017), RAD21 (Dowen et al., 2013), NIPBL (Kagey et al., 2010), and H3K27ac (Creyghton et al., 2010). CTCF orientations are indicated below CTCF track (red, forward; blue, reverse; black lines, no orientation detected). SE, *Prdm14* super enhancer (Whyte et al., 2013). See also Figure S1.

(legend continued on next page)

could act as cohesin loading platforms that help regulate cell type-specific gene expression. In support of this hypothesis, loop domains spanning SEs and NIPBL-bound sites are the fastest to recover when cohesin is reintroduced after its depletion (Rao et al., 2017).

Whereas most of our understanding of how enhancers, genes, and insulators interact to coordinate transcriptional regulation comes from genome-wide correlative studies, relatively few studies have tested and deciphered their local interplay at individual loci (Hay et al., 2016; Lupiáñez et al., 2015; Symmons et al., 2016). In this work, we focused on the 85 kb *Prdm14* domain in mouse embryonic stem cells (ESCs), which contains a SE and the pluripotency gene *Prdm14*, flanked by strongly interacting CTCF boundaries. Previous work showed that deletion of the domain's upstream CTCF boundary leads to prominent upregulation of the gene in the upstream adjacent domain, *Slco5a1* (Dowen et al., 2014). To examine how the domain boundary and surrounding DNA elements coordinate looping and transcription in these two domains, we genetically dissected regulatory sequences across the *Prdm14* locus in F1 hybrid ESCs derived from a cross of 129 and castaneous mice (Jonkers et al., 2008; Splinter et al., 2011). By creating a series of 129-specific single and compound deletions and other genetic modifications, and developing allele-specific approaches that sensitively measure the transcriptional, topological, and protein recruitment consequences of our modifications, we link the trajectory of SE-deposited cohesin to the chromosomal search space of this enhancer for target genes and show how this is orchestrated by CTCF binding elements.

RESULTS

Chromatin topology of the *Prdm14* SE domain and its neighbors

To characterize the topological features of the wild-type (WT) *Prdm14* locus, we performed 4C-seq to create contact profiles of the left and right *Prdm14* domain boundaries (LB and RB, respectively), the *Prdm14* SE, and the promoter of the *Slco5a1* gene (Slc), which sits about 60 kb upstream of the *Prdm14* contact domain (Figure 1A). As described previously (Dowen et al., 2014), the LB contains two forward-oriented CTCF sites (LB1-2), while the RB has one reverse-oriented CTCF site (RB) upstream of a series of forward-facing CTCF sites (RB2-3-4). Available chromatin immunoprecipitation sequencing (ChIP-seq) data illustrate regional binding of CTCF and cohesin subunit RAD21, with the 45 kb *Prdm14* SE, composed of five evenly distributed elements termed SE1–SE5 (Hnisz et al., 2015; Figure S1), demarcated by high enrichment for H3K27ac, cohesin loader NIPBL (also known as SCC2), pluripotency factors, and Mediator subunit MED1 (Figures 1 and S1). 4C-Seq anchored at selected viewpoints (VPs) confirms that the WT boundaries interact with each other at high frequency and encompass a distinct self-interacting domain (the "*Prdm14* domain") that contains the enhancer and *Prdm14* (Figure 1A, VPs RB, LB, and SE). Instead, the *Slco5a1* promoter, which contains a forward-oriented CTCF site immediately inside the open reading frame (5′Slc), interacts mostly with upstream sites, including a forward-oriented CTCF site near its 3′ end (3′Slc), in a more diffuse domain (the "*Slco5a1* domain") (Figure 1A, VP Slc). In all four 4C profiles, interactions of relatively low frequency between the domains are appreciable. This is also seen in available Hi-C data (Figure 1A, top; Bonev et al., 2017) and suggests that the domains are not completely insulated from each other. By ChIP-qPCR, we confirmed association of CTCF and cohesin to domain boundary CTCF sites and *Slco5a1*'s 5′ and 3′ ends (Figure 1B).

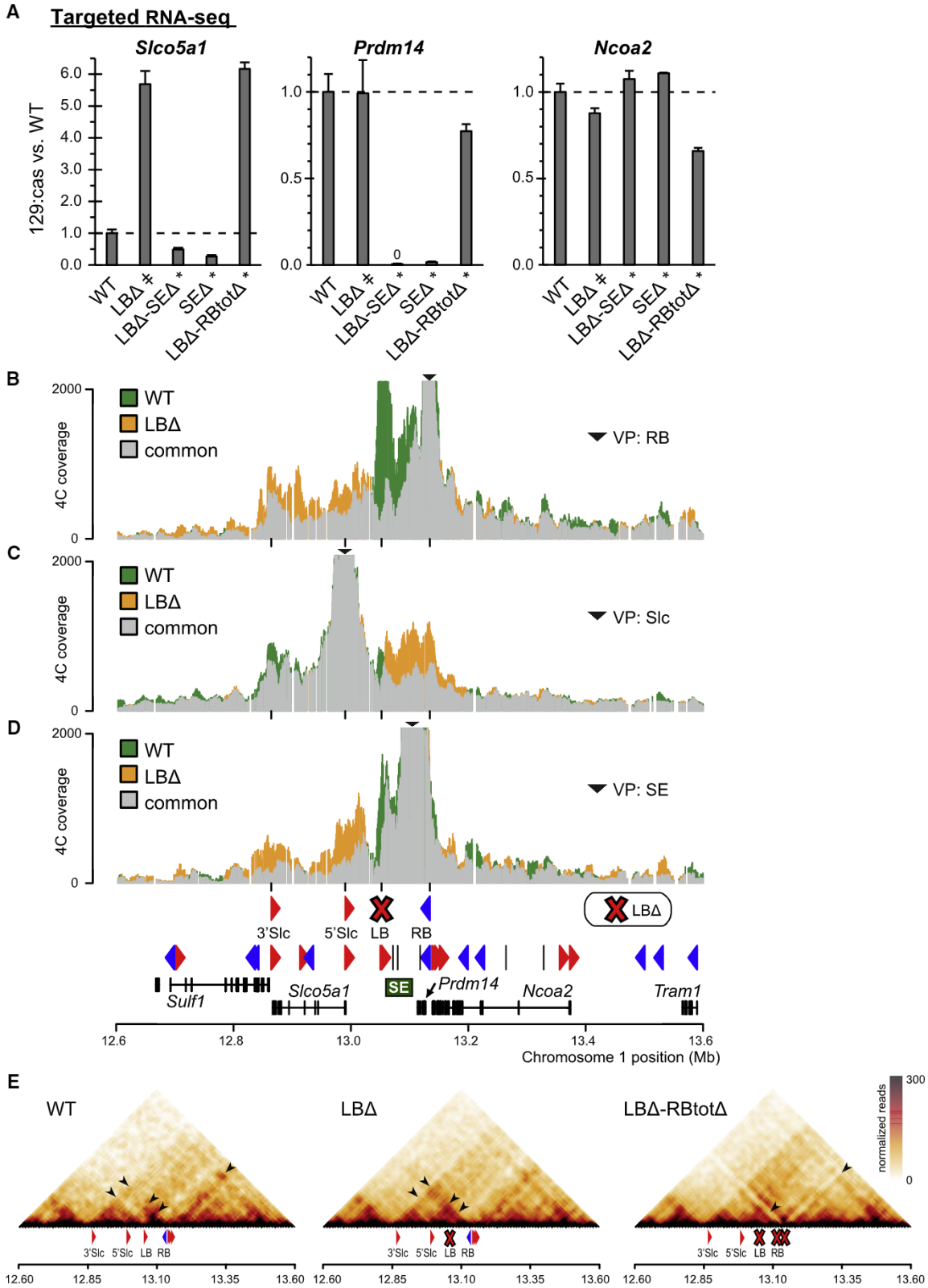
Taking advantage of the genetic variation between the two genomes in 129/castaneous ESCs, we developed allele-specific gene expression, protein recruitment (ChIP), and chromosome conformation analysis (4C-seq, capture Hi-C, and multi-contact 4C [MC-4C]) strategies to independently analyze the 129 and castaneous *Prdm14* loci (Figure 1C). When introducing CRISPR-Cas9-targeted genetic modifications, care was taken to always select clonal cell lines (clones) having the intended modification only at the 129 locus. In this manner, all clones had the same intact, WT castaneous locus, which served as an invariant, clone-intrinsic reference and enabled sensitive detection of sequence-dependent transcription and chromatin changes. To validate and generalize findings and exclude that allelic PCR biases influence results, key genetic modifications were replicated on the castaneous allele.

The *Prdm14* SE acts across its domain boundary to activate *Slco5a1* gene expression

We deleted the LB of the *Prdm14* domain on the 129 allele (LBΔ; Figure S1B) to see if we could reproduce the results of Dowen et al. (2014), in which deletion of the same two CTCF sites caused a 4.5-fold increase of *Slco5a1* expression. To monitor allelic differences in gene expression, we used next-generation sequencing to quantify, per gene, the ratio of transcripts originating from mutant 129 and WT castaneous alleles and asked whether this ratio changed relative to the 129/cas ratio in WT cells (see STAR Methods; Table S2; Figure S2A). In our LBΔ cells, we detected a 5.7-fold increase in 129 *Slco5a1* expression, while the nearby genes *Prdm14* and *Ncoa2* were not affected (Figure 2A). Upregulation of *Slco5a1* was confirmed by qRT-PCR (Figure S2B)

(B) Wild-type ChIP-qPCR enrichment for CTCF and RAD21. Neg., negative control site; pos., positive control site. Values are normalized to positive sites. Error bars represent SD for four to seven experiments, each performed in triplicate. See also Table S5.

(C) Explanation of allele-specific methods and analyses. 129/castaneous F1 hybrid ESCs were derived from the inner cell mass of a blastocyst from a cross between 129 and castaneous mice (Jonkers et al., 2008). SNPs enable identification of 129 and cas alleles. Allele-specific 4C uses inverse PCR primers with Illumina sequencing overhangs, of which the P1 primer is near the restriction site to enable reading the ligation product, and the P2 primer is near a SNP to assign each read pair to an allele. Targeted RNA-seq uses primers with Illumina sequencing overhangs, spanning an amplicon that covers a coding SNP that enables *in silico* splitting and counting of allelic gene products. Similarly, targeted ChIP-seq uses primers with Illumina overhangs, spanning an amplicon that covers the site of interest and a SNP that enables *in silico* splitting and counting allelic binding events by protein of interest. Allele-specific MC-4C: inverse PCR primers exclusively amplify ligation products of the 129 allele. Allele-specific capture Hi-C uses capture probes designed against the reference genome, with sequenced reads split *in silico*. See also STAR Methods.



(legend on next page)

and is consistent with the aforementioned research (Dowen et al., 2014), in which it was assumed that boundary deletion enabled *Slco5a1* to hijack the *Prdm14* SE. To directly test this, we investigated whether the LB deletion (LBΔ) indeed fused the *Prdm14* and *Slco5a1* domains: allele-specific 4C-seq applied to the RB in these cells revealed that the *Prdm14* domain-spanning loop disappeared, intra-domain contacts decreased, and contacts across the entire *Slco5a1* domain increased (Figure 2B). A similar re-distribution of contacts between both domains was seen when we profiled 129 contacts of the *Slco5a1* promoter (Figure 2C) and *Prdm14* enhancer (Figure 2D). These contact profiles further uncovered that in the absence of the left *Prdm14* domain boundary, contacts between *Slco5a1* and the SE increased (Figures 2C and 2D). Collectively, our high-resolution conformation capture data demonstrate that deletion of the LB fuses the two flanking domains, enabling *Slco5a1* to contact the SE more frequently. To corroborate our 4C-seq data, we performed allele-specific capture Hi-C (see STAR Methods) in WT and LBΔ cells. In agreement with our 4C-seq results, we observed merging of the *Slco5a1* and *Prdm14* domains, specifically on the 129 allele having the LB deletion (Figures 2E and S7). To further validate these findings, an independent clonal cell line was created, now with the LB deletion on the castaneous allele and an intact 129 locus. This recapitulated the results, with a 5.8-fold upregulation of *Slco5a1* (Figure S3A), reduced insulation between the domains, and increased contacts between *Slco5a1* and the *Prdm14* SE on the castaneous allele (Figure S3B), thus confirming the phenotype and excluding that results originate from allelic biases.

To functionally test whether the SE is responsible for increased *Slco5a1* expression, we further modified the 129 LBΔ cells by also deleting the 129 SE, creating two 129 LBΔ-SEΔ clones. *Prdm14* expression was completely abolished (Figure 2A). Expression of *Slco5a1* was reduced nearly 12-fold compared with 129 LBΔ cells and 2-fold compared with WT. This confirmed that the SE regulates *Slco5a1* expression in LBΔ cells and suggested that also in WT cells, the SE contributes to expression of *Slco5a1* in the neighboring domain. To directly test this, we deleted the 129 SE in two 129/cas WT clones (SEΔ) and found that this reduced *Slco5a1* expression by 70% (Figure 2A). Thus, despite being flanked by domain boundaries that frequently interact, the enhancer is able to increase expression of a gene located in a neighboring domain.

Expression of *Ncoa2*, a gene in the immediate downstream domain, that was not appreciably contacted by the SE (Figures 1A and 2D) was not affected by the LB deletion (LBΔ), SE deletion (SEΔ), or the combined deletion of both (LBΔ-SEΔ) (Figure 2A). Although its expression was somewhat reduced when both boundaries were deleted in their entirety (LBΔ-RBtotΔ), in the latter clones (two replicates) we also deleted the 3′ end of *Ncoa2* in order to delete the CTCF sites, potentially affecting expression (Figures 2A and S1). Thus, unlike *Slco5a1*, expression of *Ncoa2* is apparently not influenced by the *Prdm14* SE. As expected, complete deletion of both boundaries led to collapse of overall *Prdm14* domain structure, as shown by allele-specific capture Hi-C (Figures 2E and S7).

Slco5a1's promoter-associated CTCF site has a minor effect on transcriptional activation by the *Prdm14* SE

We were interested in the role of the CTCF site positioned near the start of the *Slco5a1* gene, in its first exon, which we had found engaged in looping with boundaries and the *Prdm14* SE (5′Slc; Figures 1A and 2B–2D) and which we suspected could be involved in this gene's regulation. We therefore created two independent clones carrying six mutated nucleotides within the CTCF core motif on the 129 allele, preserving the translation sequence (LBΔ-mutC; see STAR Methods). Applying CTCF and cohesin (RAD21) ChIP-qPCR on these heterozygous cells and WT cells demonstrated a 50% reduction in binding at the *Slco5a1* promoter, indicating that the disrupted binding motif of the 129 allele no longer enabled CTCF recruitment (Figure 3A). Without CTCF binding to the promoter, we observed that expression of *Slco5a1* was reduced by about 30% compared with its parent cell line, LBΔ, but still strongly elevated compared with WT cells (Figures 3B). Long-range gene activation of the *Slco5a1* gene by the *Prdm14* SE therefore benefits from, but does not solely depend on, CTCF recruitment to its promoter. Allele-specific 4C-seq confirmed that the *Slco5a1* promoter lacking CTCF was not hampered in its interactions with the SE. Instead, compared with its parent clone LBΔ, it mainly lost contacts with the CTCF-bound RB (Figures 3C and S4A).

The *Prdm14* SE forms a hub that can simultaneously contact two genes

We found that both the *Slco5a1* and *Prdm14* genes relied on the *Prdm14* SE for their WT transcription levels, yet *Prdm14* expression was not affected when *Slco5a1* increased its interaction with the SE to activate its expression in LBΔ cells (Figure 2A). This demonstrated that *Slco5a1* has limited capacity to functionally compete with *Prdm14* for the SE, which is perhaps not surprising given that *Slco5a1* is expressed at a much lower level than *Prdm14* (average 0.5 versus 7.5 fragments per kilobase of transcript per million mapped reads [FPKM] in WT; Bonev et al., 2017). We asked whether, oppositely, *Prdm14* influenced activated *Slco5a1* gene expression. To examine this, we first deleted the *Prdm14* promoter on the WT allele (PΔ). This

Figure 2. The *Prdm14* super enhancer acts across its domain boundary to activate *Slco5a1* gene expression

(A) Targeted RNA-seq in WT and mutant clones, plotted per gene as 129/cas expression ratios normalized to the ratio in WT 129/cas cells (see STAR Methods). A ratio of 1 (dashed line) indicates no difference versus normalized WT ratio. Error bars represent SD of technical (double dagger) or biological (asterisk) replicates per genotype, per gene (WT, n = 15 or 16; mutants, n = 2). See also Figure S2 and Table S2.

(B–D) 4C profile overlays comparing chromatin contacts for (B) the right boundary (RB), (C) the *Slco5a1* promoter, and (D) the super enhancer (SE) on the mutant 129 allele with left boundary deletion (LBΔ) versus those on the WT 129 allele, in mutant and WT cells; common (no difference) interactions in gray. y axis, 4C coverage per 1 million normalized reads. Bottom: CTCF orientations, super enhancer (SE), and genes.

(E) Allele-specific capture Hi-C in WT, LBΔ, and LBΔ-RBtotΔ alleles. Arrowheads highlight merging of the *Prdm14* and *Slco5a1* domains in LBΔ cells and of the *Prdm14* domain with both flanking domains in LBΔ-RBtotΔ cells. See also Figure S7.

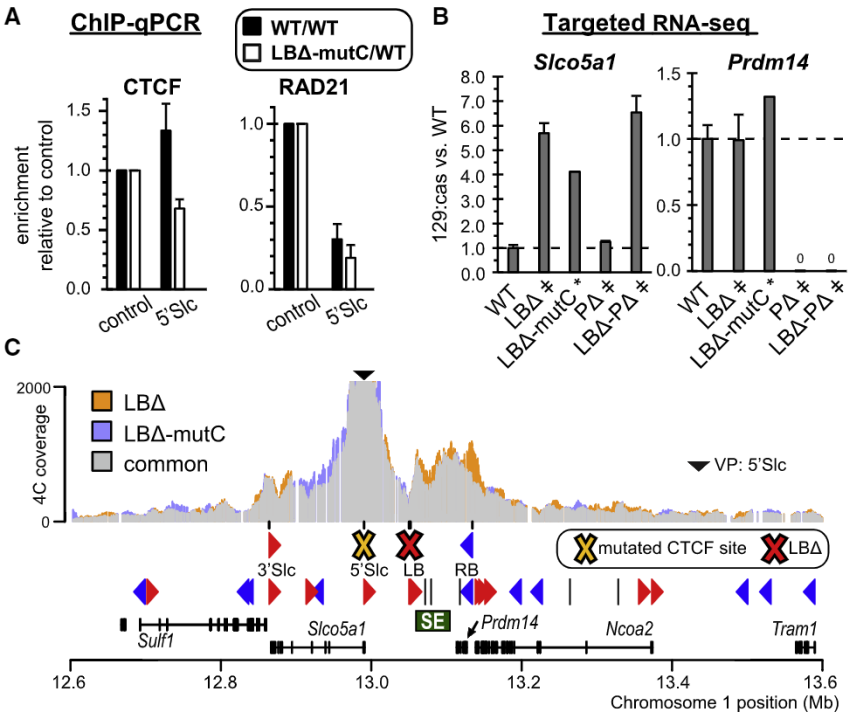


Figure 3. *Slco5a1*'s promoter-associated CTCF site has a minor effect on transcriptional activation by the *Prdm14* super enhancer

(A) Abolished CTCF binding to the 129 *Slco5a1* promoter in LBΔ-mutC cells. ChIP-qPCR shows efficient *Slco5a1* promoter binding of CTCF in WT cells, which is 50% reduced in LBΔ-mutC cells having a 129 allele with disrupted CTCF-binding motif. RAD21 shows a similar 50% reduction in LBΔ-mutC cells. Values are normalized to the control site. Error bars represent SD for two biological clones; six or seven WT replicates; each performed in triplicate. See also Table S5.

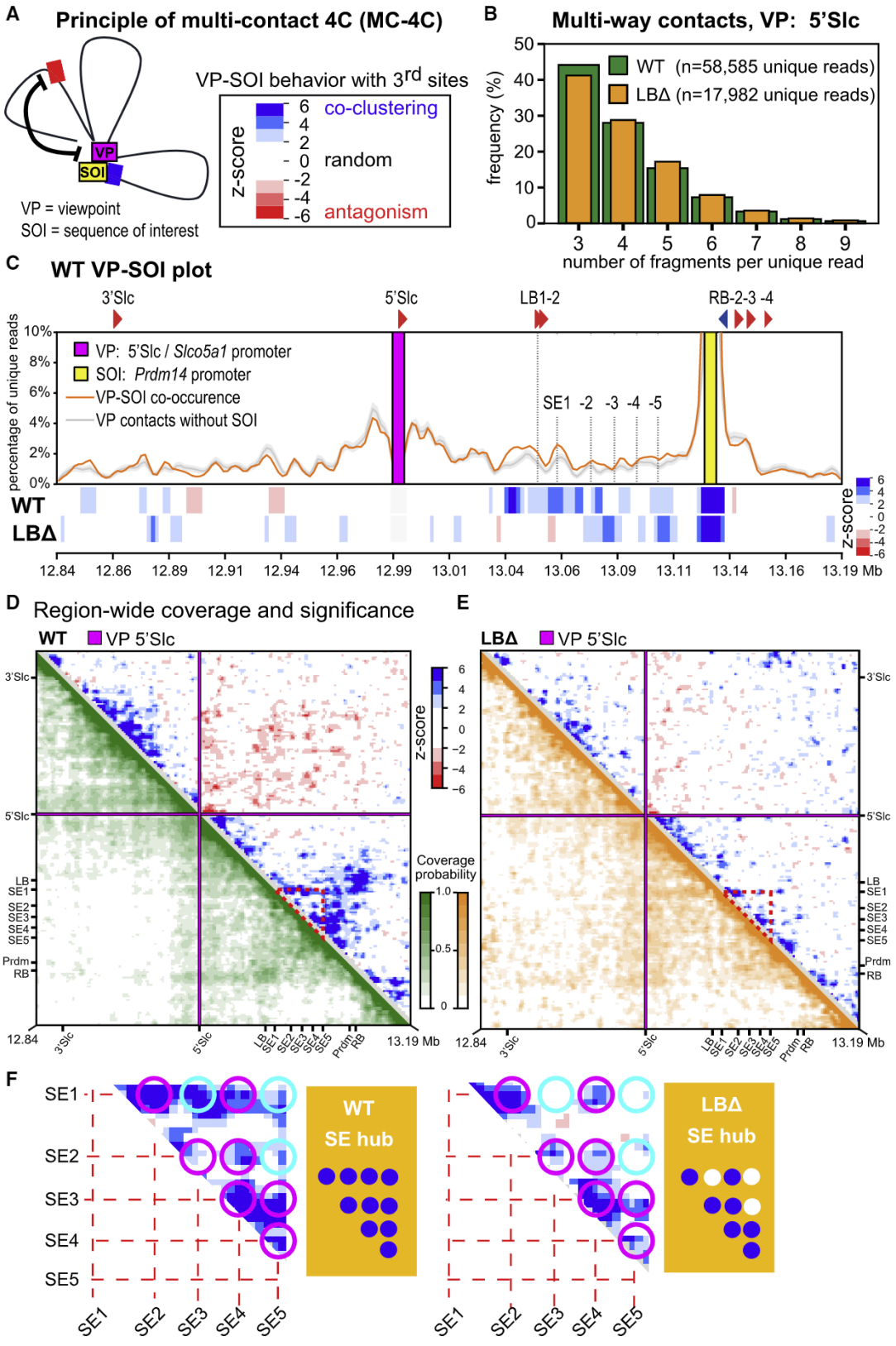
(B) *Slco5a1* 129/cas gene expression ratios in 129 mutant clones, normalized to the ratio in WT cells. A ratio of 1 (dashed line) indicates no difference versus WT ratio. Error bars represent SD of technical (double dagger) or biological (asterisk) replicates per genotype, per gene (WT, n = 15 or 16; mutants, n = 2). See also Figure S2 and Table S2.

(C) 4C profile overlays comparing chromatin contacts for the *Slco5a1* promoter on the mutant 129 allele having the left boundary deletion and mutated *Slco5a1* promoter-associated CTCF site (LBΔ-mutC), compared with those on the mutant 129 allele with only the left boundary deletion, as measured separately in each cell line. Common (no difference) interactions in gray. y axis, 4C coverage per 1 million normalized reads. Bottom: CTCF orientations, super enhancer (SE), and genes.

abolished *Prdm14* expression and only marginally increased *Slco5a1* expression (Figures 3B). Deleting the *Prdm14* promoter in the context of the 129 LBΔ allele (LBΔ-PΔ) likewise abolished *Prdm14* expression, while *Slco5a1* increased to an average 6.5-fold versus WT, an additional 15% compared with the 5.7-fold activation seen in the LBΔ background (Figure 3B). These results indicated limited, if any, functional competition between the genes for transcriptional activation by the enhancer.

To understand whether there is physical competition between the two genes for the SE, and to study the higher order topology of the locus, we performed MC-4C (Allahyar et al., 2018; Vermeulen et al., 2020). MC-4C is a 3C-based method that uses long-read sequencing (Oxford Nanopore) to analyze tens of thousands of allelic multi-way contacts (mostly three- to five-way interactions) centered on a genomic site of interest (the VP). Multi-way contact analysis enables one to assess the statistical significance of co-occurrence frequencies in terms of *Z* scores, to distinguish whether, on the same single allele, two genomic sites interact with the VP in a competitive (antagonistic), cooperative (clustering), or random fashion (Figure 4A). We designed an allele-specific PCR strategy (see STAR Methods) to exclusively amplify and analyze 58,585 and 17,982 multi-way contacts formed by the 129 *Slco5a1* gene promoter in WT and 129 LBΔ cells, respectively (Figure 4B). The orange line in Figure 4C shows, for all sequences across the WT locus, how frequently they are found together with the *Slco5a1* promoter when it is interacting with a sequence of interest (SOI; here the *Prdm14* promoter), compared with when it is not interacting with the SOI (gray line). A comparison between these frequencies allows one to calculate whether the sequences cooperate (dark blue *Z* scores) or compete (dark red *Z* scores) with the *Prdm14* promoter (SOI) for being in contact with the *Slco5a1* promoter (VP). This so-called VP-SOI plot reveals that sequences across the *Prdm14* SE form preferred three-way contacts with the *Prdm14* promoter and the *Slco5a1* promoter. Reciprocally, we also observed that when the *Slco5a1* promoter was in contact with any of the individual regulatory sites of the SE, the *Prdm14* gene promoter was likely to be interacting as well (Figure S5).

To better appreciate spatial co-clustering across the locus as a whole, we plotted co-occurrence frequencies and calculated significance scores side by side in matrices for the entire 350 kb region of interest, for WT and LBΔ alleles (Figures 4D and 4E). In WT cells, we found a clear separation between the *Slco5a1* and *Prdm14* domains: the *Slco5a1* promoter either co-localized with sequences of its own domain, or simultaneously contacted multiple sequences in the *Prdm14* domain, but was unlikely to have sequences of both domains in a given micro-topology (Figure 4D). This separation of contacts between domains was clearly diminished in LBΔ cells (Figure 4E). Additionally, as suggested by individual VP-SOI plots (Figure S5), preferential clustering among the individual regulatory elements of the SE was also evident in the WT association matrix (red triangle, Figure 4D; enlargement, Figure 4F) but was less pronounced on the LBΔ allele (Figures 4E and 4F). Collectively, these results suggest that the individual regulatory sites of the *Prdm14* SE aggregate to form an enhancer hub that can physically accommodate two genes at the same time, as seen before for the β-globin SE (Allahyar et al., 2018). The data also suggest that the physical stability of the enhancer hub benefits from being situated between interacting CTCF boundaries.



(legend on next page)

SE insulation is not controlled by looping of its flanking boundaries

We wondered whether we could neutralize the LB to phenocopy the LB deletion (Figure 2) by disrupting its loop with the RB. For this, we deleted the reverse-facing CTCF site anchoring this loop at the RB (RBΔ) and, independently, inverted the LB sequence (LB-inv). Applying 4C-seq to the LB in RBΔ cells confirmed that this deletion caused major disruption, but not complete abolishment, of its chromatin loop (Figure 5A). Similarly, 4C-seq applied to LB-inv cells confirmed that the right *Prdm14* boundary (RB) no longer formed a loop with the inverted LB (Figure 5B). In contrast to LBΔ cells though, *Slco5a1* was not upregulated in RBΔ or LB-inv cells. In fact, in RBΔ, *Slco5a1* expression was downregulated rather than upregulated, as was expression of *Prdm14* (Figures 5C and 5D). A topological underpinning for the downregulation of *Slco5a1* may come from examining its chromatin contacts in RBΔ cells, which showed that fewer were made with the *Prdm14* domain and the SE, while contacts were promoted with sequences downstream of the deleted RB (Figure 5E). In LB-inv cells, the inverted LB made strong new contacts across the entire *Slco5a1* domain, including the forward-oriented CTCF sites at the gene's 5′ and 3′ ends (Figure 5B). 4C-seq from the 5′ end of *Slco5a1* confirmed strong insulation from the SE in LB-inv cells (Figure 5F). In both RBΔ and LB-inv, 4C-seq from the SE showed only a minor re-distribution of contacts (Figures S4B and S4C). These findings are in stark contrast with the new contacts formed between the *Slco5a1* promoter and the SE upon deletion of the LB (Figures 2C and 2D). Allele-specific capture Hi-C on LB-inv cells supported the conformational changes observed using 4C-seq (Figures 5G and S7). To validate our LB-inv results, we generated a castaneous LB-inv cell line with an intact 129 allele. Also here we found no upregulation of *Slco5a1* expression upon boundary inversion (Figure S6A), and we confirmed that the inverted sites no longer loop downstream to the RB but upstream to the CTCF sites flanking the *Slco5a1* gene (Figure S6B). We conclude that the LB retained its capacity to insulate the SE even when we disrupted its looping ability with the RB. We further observed that the reverse-oriented CTCF site at the RB appeared to support the SE in its regulatory activity within and beyond the domain (Figures 5C and 5D).

The SE contributes cohesin to surrounding CTCF sites

We finally investigated whether recruitment of CTCF and cohesin was affected by our genetic modifications, perhaps in a way that could explain their transcriptional and topological consequences. Analogous to how we calculated gene expression ratios, we performed targeted ChIP-seq and counted the number of times each selected 129- or cas-specific SNP was detected in our NGS reads. We then took the ratio of 129 to cas SNP counts for each site and normalized these to their respective 129/cas average ratios in replicate WT/WT cells (see STAR Methods; Table S4). Reassuringly, the WT 129/cas ratios were always very close to 50:50, across all experiments and for all sites, with the exception of RB4, which consistently recruited CTCF and cohesin more efficiently to the 129 allele (Figure S2D). This, we found, was apparently caused by two base substitutions weakening the CTCF consensus motif on the cas allele (Figure S2E).

We first asked whether deletion of one of the boundaries had consequences for CTCF or cohesin accumulation at the opposite boundary. Both in LBΔ and RBΔ alleles, CTCF and RAD21 recruitment to the opposite intact boundary was unaffected (Figures 6A and 6B). The inverted LB sites (LB-inv) recruited normal levels of CTCF and perhaps slightly reduced levels of cohesin. Overall, therefore, the degree of cohesin accumulation at CTCF binding sites seems not much related to their looping engagement with other CTCF sites (de Wit et al., 2015). In fact, CTCF and cohesin levels at both boundaries were relatively stable across all mutant alleles, except for the alleles that lacked the SE (SEΔ and LBΔ-SEΔ). Specifically in these mutants, RAD21 levels clearly dropped at the flanking boundary CTCF sites LB1, LB2, and RB. This strongly suggests that the *Prdm14* SE has an important role in cohesin recruitment and deposition at its surrounding CTCF binding sites. This idea received further support from the observation that specifically in LBΔ alleles with an intact SE (LBΔ, LBΔ-mutC, and LBΔ-PΔ), RAD21 levels increased at the SE's next most proximal, intact upstream CTCF binding sites, the *Slco5a1* promoter (5′Slc, except when mutated) and 3′ end (3′Slc). These observations were recapitulated in our LBΔ and LB-inv cell lines with modifications on their respective castaneous alleles (Figures S3C and S6C). In compound LBΔ-SEΔ alleles having also the SE deleted, RAD21 levels at these sites returned to WT levels (Figure 6, 3′Slc and 5′Slc). We conclude that the immediately surrounding CTCF sites are able to accrue cohesin whose loading originates at the SE. This phenotype, in which disruption of a given DNA binding site results in protein accumulation specifically at the next *cis*-linked binding site (tens of kilobases away), provides strong *in vivo* support for cohesin traversing the chromatin fiber and contributes supporting evidence for the loop extrusion model.

Figure 4. The *Prdm14* super enhancer forms a hub that can simultaneously contact two genes

(A) MC-4C estimates a background (expected) profile to identify preferential (co-clustering), random, or competitive (antagonistic) interactions between a viewpoint (VP), sequence of interest (SOI), and third sites across the region of interest. The significance of interactions between a VP and SOI with third sites are assessed in terms of *Z* scores.

(B) Distribution of multi-way contacts detected for VP 5′Slc in WT and LBΔ alleles. The number of interacting DNA fragments includes the VP. Fragments = restriction fragments. Note that the distributions are similar for the two experiments despite the different number of unique reads.

(C) When the *Slco5a1* and *Prdm14* promoters are in contact, SE elements are also likely to cluster. Top: VP-SOI plot for VP 5′Slc (magenta) and SOI Prdm14 (yellow) on the WT allele. Gray line, background contact profile excluding the SOI, ±1 SD (gray shading). Orange line, co-occurring sequences when VP and SOI interact. Bottom: *Z* scores across the region for WT and LBΔ alleles. See also Figure S5.

(D and E) Plots show normalized region-wide MC-4C results for (D) WT and (E) LBΔ alleles using the *Slco5a1* promoter viewpoint. Per plot, lower left shows the measured coverage (green, WT; gold, LBΔ), and upper right shows the calculated significance in terms of *Z* scores. Red dashed triangle highlights SE region.

(F) Enlargement of SE region. Note that individual interactions between SE elements 1–5 (circles) are relatively depleted in LBΔ versus WT, summarized in the cartoons (white dots, loss of clustering).

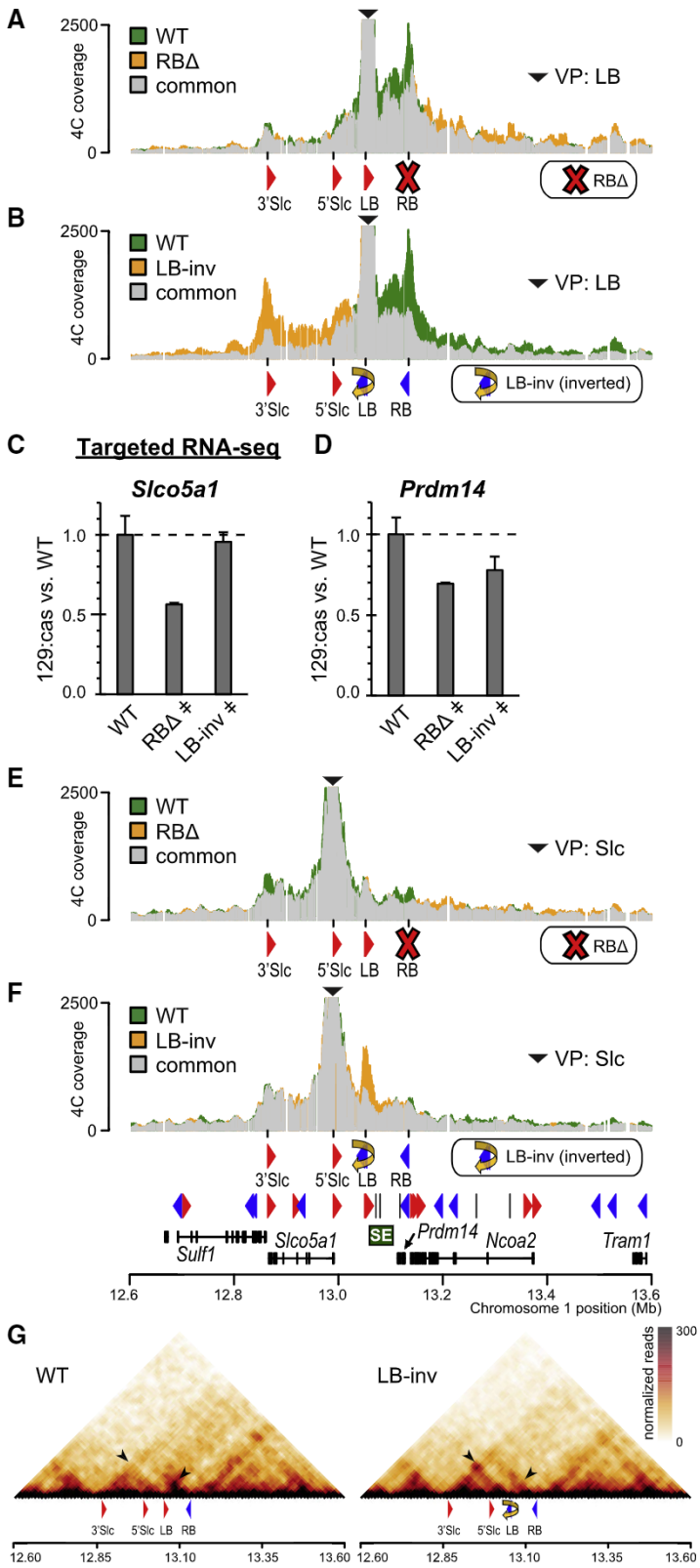


Figure 5. Super enhancer insulation is not controlled by looping of the flanking boundaries

(A and B) 4C profile overlays comparing chromatin contacts of the left boundary (LB) on the mutant 129 allele having (A) the convergent right boundary CTCF site deleted or (B) the left boundary CTCF sites inverted, compared with the WT 129 allele. Common (no difference) interactions in gray. y axis, 4C coverage per 1 million normalized reads. Bottom: CTCF orientations, super enhancer (SE), and genes.

(C) *Slco5a1* 129/cas gene expression ratios in 129 mutant clones, normalized to the ratio in WT cells. A ratio of 1 (dashed line) indicates no difference versus WT ratio. Error bars represent SD of technical (double dagger) replicates per genotype, per gene (WT, n = 15 or 16; mutants, n = 2). See also Figure S2 and Table S2.

(D) As in (C) but for *Prdm14*.

(E and F) 4C profile overlays of *Slco5a1* promoter contacts on the mutant 129 allele in (D) RBΔ cells and (E) LB-inv cells, compared with the WT 129 allele (green) in WT cells. Data are presented as in (A). See also Figure S4.

(G) Allele-specific capture Hi-C on WT and LB-inv alleles. Arrowheads highlight the new loop formed by the inverted LB CTCF sites with upstream CTCF sites in the *Slco5a1* domain and loss of looping across the *Prdm14* domain. See also Figure S7.

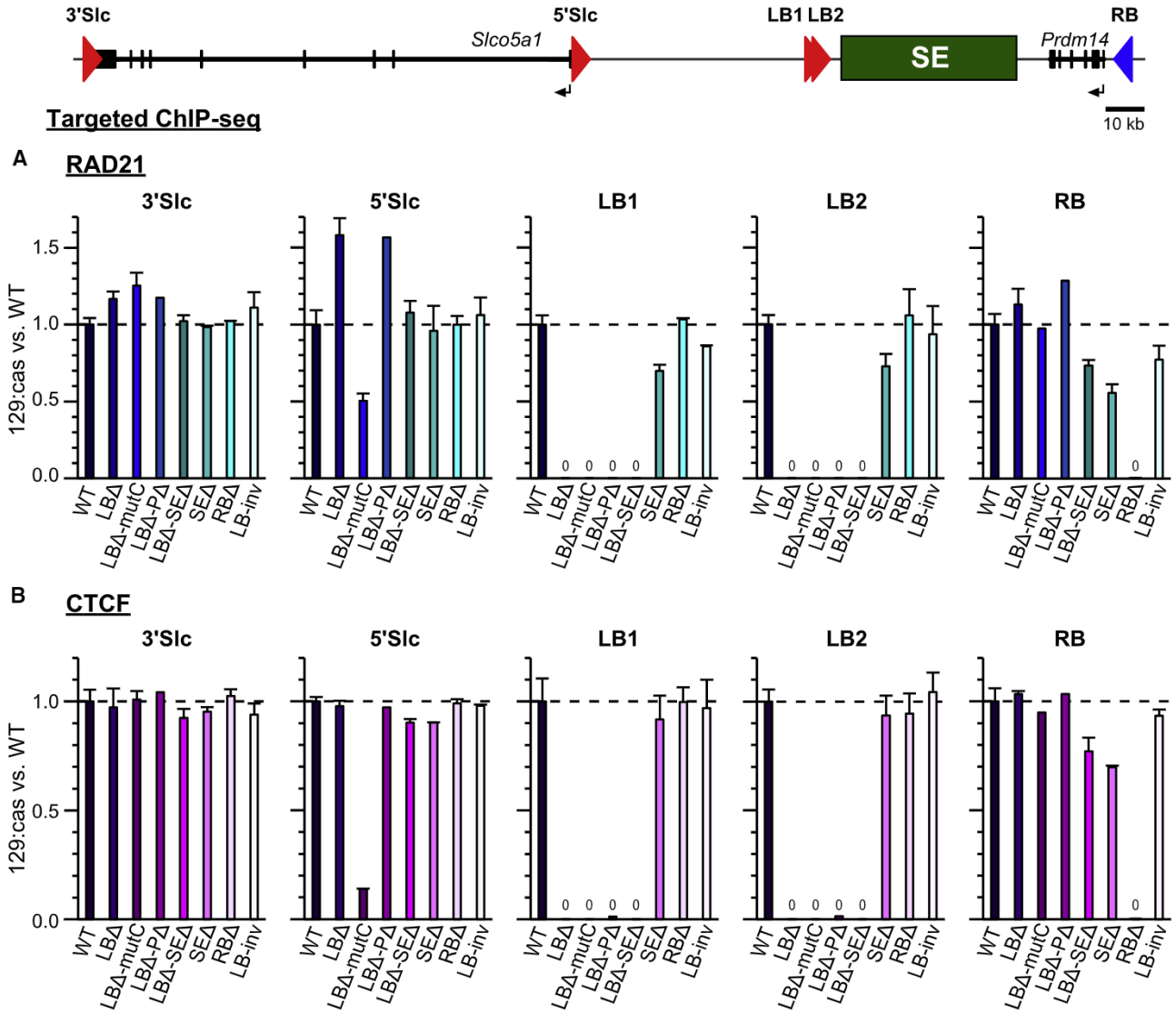


Figure 6. The super enhancer contributes cohesin to surrounding CTCF sites

(A) Targeted ChIP-seq results for RAD21 at indicated sites. 129/cas enrichment ratios in mutant clones, normalized to the ratio in WT cells. A ratio of 1 (dashed line) indicates no difference versus WT. Error bars represent SD (WT, n = 7–9; mutants, n = 2). See STAR Methods and Table S4.

(B) As in (A) but for CTCF.

Last, some surprising observations were made regarding oppositely oriented CTCF sites immediately flanking the SE. First, we found that RAD21 accumulation at 5′Slc, reproducibly seen in LBΔ alleles and dependent on the SE, was not observed in LB-inv alleles, having inverted LB CTCF sites engaged in upstream-oriented loops (Figures 6A and S6C). This suggests that cohesin loaded at the SE does not bypass the inverted CTCF sites. Second, in RBΔ alleles, in which the single RB CTCF site facing the SE was deleted (RBΔ), no increased cohesin deposition was observed on any of the immediately following, oppositely oriented CTCF sites (RB-2-3-4; Figure S2C). Collectively, this suggests that when facing away from the incoming cohesin machinery, chromatin-bound CTCF molecules that are engaged in looping block extrusion paths, yet without stalling cohesin at the CTCF site.

DISCUSSION

Our studies uncover the complex interplay between a SE and its flanking boundary CTCF sites to coordinate gene expression, cohesin deposition, and loop formation and highlight the diverse context- and orientation-dependent roles of CTCF in these processes.

An unexpected observation was that insulation by a strong boundary is not absolute. Even though the LB contains two strong CTCF binding sites, frequently forms a chromatin loop with the opposite boundary encompassing the SE, and clearly compromises functional and physical crosstalk between the SE and *Slco5a1* in the neighboring domain (Dowen et al., 2014 and this work), it still allows the SE to

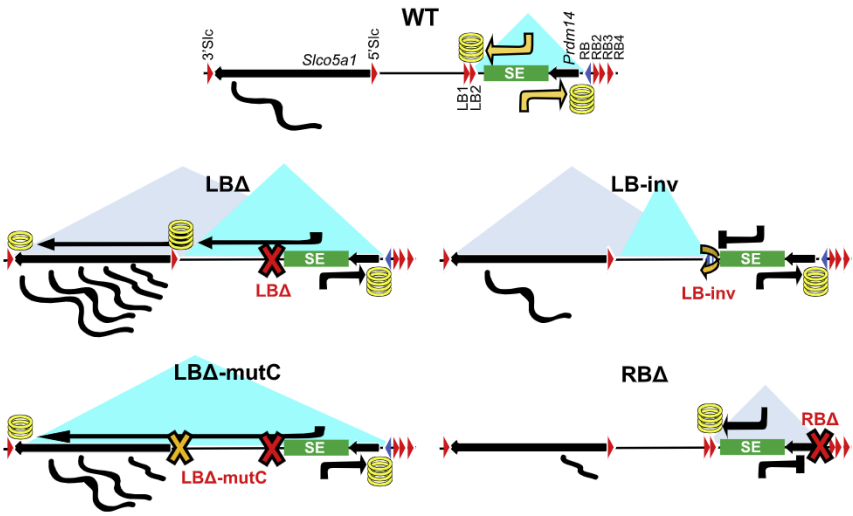


Figure 7. Depiction of loop extrusion trajectories on the basis of chromatin conformation, gene expression, and cohesin binding

In LBΔ, cohesin extrudes from the SE into the neighboring domain, accumulating at 5′ and 3′Slc sites, merging the domains and activating *Slco5a1*. In LBΔ-mutC, cohesin does not pause at 5′Slc, but some accumulates at 3′Slc; *Slco5a1* expression is slightly reduced. In LB-inv, new insulating loops are formed between the inverted LB sites and the newly convergent CTCF sites. In RBΔ, looping to the LB is reduced, as is regulatory activity of the SE. See Discussion.

account for 70% of *Slco5a1*'s expression in the WT configuration. This finding shows that although DNA contact profiles are useful in determining gene-regulatory landscapes, one must avoid being too strict when interpreting the functional consequences of conformational features such as insulating domains and boundaries. Apparently, the rare contacts of the enhancer with sites outside its own domain are sufficient to enhance *Slco5a1* expression.

We find that the SE is responsible for transcriptional activation of two genes, but also for the recruitment and deposition of cohesin on flanking CTCF sites. Figure 7 summarizes our findings and illustrates how the different CTCF sites in this locus coordinate these two activities of the SE. Without the two convergently oriented CTCF sites of the LB (LBΔ), cohesin traversing from the SE is no longer halted at this location but proceeds extruding to the next convergent CTCF site, just inside the *Slco5a1* gene body. Contacts between the *Slco5a1* gene and the SE are stimulated and simultaneously, *Slco5a1* is upregulated. When we then also mutate the CTCF site at the *Slco5a1* promoter (LBΔ-mutC), cohesin continues migrating and stabilizes more frequently at the next convergently oriented CTCF site, at the 3′ end of *Slco5a1*. In this same mutant, the *Slco5a1* gene remains strongly upregulated, demonstrating that enhancer-mediated gene activation does not strictly depend on paused loop extrusion at the CTCF site near the gene promoter. Given that even without CTCF binding to the *Slco5a1* promoter, it continues to form frequent contacts with the enhancer, other factors, such as the tissue-specific transcription factors OCT4, SOX2, and NANOG, are likely primarily responsible for this enhancer-promoter loop.

Both inversion of the LB CTCF sites (LB-inv) and deletion of the single convergently oriented CTCF site of the RB (RBΔ) disrupts looping between the *Prdm14* domain boundaries, yet without causing upregulation of *Slco5a1*. Thus, insulation of the SE is not controlled by looping between its flanking boundaries per se. The inverted CTCF sites, now engaged in looping with the *Slco5a1* domain, likely prevent cohesin loaded at the enhancer from bypassing and extruding to *Slco5a1*'s 5′ and 3′ CTCF sites. This may imply that loop-engaged, wrongly oriented CTCF sites halt extruding cohesin complexes traveling from the SE: they may immobilize them, promote their release from chromatin, and/or alter their extrusion trajectories. To investigate the first option, we focused on the RB, where a convergently oriented CTCF site is immediately followed by three oppositely oriented CTCF sites. When the convergently oriented CTCF site was deleted, no accumulation of cohesin was observed at any of the subsequent CTCF sites. Collectively, we suggest that extruding cohesin complexes do not traverse and are not stabilized at oppositely oriented, loop-engaged CTCF molecules but are either released or remain associated to the chromatin domain that recruited them.

Combined with literature on cohesin loop extrusion (Banigan et al., 2020; Davidson et al., 2019; Fudenberg et al., 2017; Ganji et al., 2018; Kim et al., 2020; Li et al., 2020; Pugacheva et al., 2020; Vian et al., 2018), including a recent study that provides genome-wide evidence for tissue-specific transcription factors converting enhancers into preferred sites for cohesin loading (Liu et al., 2021), our studies support a loop extrusion model in which cohesin complexes are preferentially loaded at strong tissue-specific enhancers (Dowen et al., 2013; Hnisz et al., 2013; Liu et al., 2021). We infer that loaded cohesin complexes then slide and extrude unanchored in both directions, possibly exposed to WAPL for release from chromatin (Haarhuis et al., 2017). When encountering a convergently bound CTCF molecule, the cohesin complex's conserved essential surface (CES) domain will interact with the CTCF's N-terminal F/YXF motif (Li et al., 2020), which stabilizes cohesin on the DNA and protects it against WAPL release, giving a prominent signal in ChIP experiments (Li et al., 2020). We found that the convergently oriented CTCF site in the RB supported *Slco5a1* and *Prdm14* expression, suggesting that a DNA-bound CTCF molecule can stimulate functional enhancer-promoter interactions if it anchors a cohesin-mediated DNA extrusion trajectory that encompasses both elements. When, however, the CES encounters a divergently bound, loop-engaged CTCF molecule (as in LB-inv and RBΔ cells), extrusion ceases, possibly to accumulate unanchored loops at the near side of the boundary, but with cohesin's CES unprotected and thus accessible to WAPL for release. The role of tissue-specific transcription factors may be dual: not only to facilitate local cohesin recruitment but likely also to help stabilize loops formed by cohesin between enhancers and promoters (Deng et al., 2012).

Limitations of the study

The careful dissection of an individual genetic locus can complement genome-wide studies, as it enables experimentally testing the relevance of observed correlations, for example in the interplay between regulatory and architectural elements. However, findings made at a given locus may not necessarily hold true at another locus.

The field refers to distinct patterns of chromatin contacts between sequences seen in Hi-C contact matrices as TADs, sub-TADs, or structural domains, but there is no objective measure to make an absolute distinction between each of these domain classes. Our classification of the "*Prdm14* domain" is subjective and arbitrary: Hi-C shows that it spans a distinct self-interacting chromatin region, flanked by convergent CTCF sites. We use the term "*Slco5a1* domain" even more arbitrarily, as a means to refer to the region upstream of the "*Prdm14* domain."

Our conclusion that the *Prdm14* SE can physically accommodate two genes at the same time is based exclusively on MC-4C. Ideally, this would be orthogonally validated by fluorescence *in situ* hybridization (FISH). As the *Prdm14* gene is only 20 kb away from the SE (itself 45 kb), and also *Slco5a1* is only 70 kb from the SE, we here had no means to distinguish between different topologies by FISH.

We developed targeted ChIP-seq and RNA sequencing (RNA-seq) methods to carefully measure the *cis*-regulatory impact of modifications on the 129 allele in cells having a reference WT castaneous allele. These methods require simultaneous PCR amplification with one primer set of two amplicons with allele-distinguishing SNPs, followed by next-generation sequencing and SNP counting to determine allelic ratios. Although a negligible impact on PCR amplification efficiency is expected from a single SNP, or even a few SNPs, located in between the priming ends of an amplicon, we excluded the possibility of PCR bias by normalizing allelic ratios to many replicates of WT cells and by finding the same results in cells having the opposite allele targeted (i.e., carrying a mutant castaneous and WT 129 allele).

On the basis of our observation that cohesin levels consistently accumulated at the next convergent CTCF site when we deleted those immediately flanking the *Prdm14* SE (in LBΔ, LBΔ-mutC, and LBΔ-PΔ cells), and that this was abolished when we deleted it (in SEΔ and LBΔ-SEΔ cells), we concluded that the SE serves as a cohesin recruitment platform. Although in agreement with recent literature (Liu et al., 2021), we also considered the possibility that the drop in cohesin levels at flanking CTCF sites could be explained by the fact that the 45 kb deletion, required to remove the SE, simply reduced the genomic interval probed by cohesin for association. However, the reverse-oriented RB CTCF site helped distinguish between these options: its cohesin level was resistant to increased changes in size of the *Prdm14* domain when the LB was deleted, but dropped when the enhancer was deleted (in SEΔ and LBΔ-SEΔ cells). The presence or absence of the SE therefore appeared to be the dominant feature associated with altered cohesin accumulation at flanking convergent CTCF sites.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2021.06.008.

ACKNOWLEDGMENTS

We thank Geert Geeven for help with 4C analysis and cHi-C bait design, Valerio Bianchi for help with gene expression and ChIP analyses, Julen Mendieta-Esteban and Marc Marti-Renom at CNAG-CRG Barcelona for help with MC-4C analysis, Robin van der Weide at NKI for help with insulation scores, and de Laat lab members for discussions. This work is part of the Oncode Institute and was funded by a VICI grant (724.012.003) from the Netherlands Organization for Scientific Research (NWO) and a Fondation Leducq Transatlantic Network grant (14CVD01).

AUTHOR CONTRIBUTIONS

E.S.M.V. and W.d.L. conceived experiments. E.S.M.V. designed genome editing. C.V.-Q. designed and C.V.-Q. and M.J.A.M.V. performed ChIPs. Y.H. and P.H.L.K. designed and performed capture Hi-C. E.S.M.V., M.J.A.M.V., F.v.d.V., and E.C.H.U. generated and/or validated cell lines. E.S.M.V. designed and E.S.M.V. and M.J.A.M.V. performed 4C, MC-4C, and gene expression experiments. M.J.A.M.V. and A.-K.F. performed qRT-PCR experiments. P.H.L.K. and A.A. performed bioinformatic analyses and helped with experimental design. E.S.M.V., P.H.L.K., and A.A. produced figures. E.S.M.V. and W.d.L. wrote the manuscript with input from other authors.

DECLARATION OF INTERESTS

W.d.L. is a founder and shareholder of Cergentis BV.

Received: June 23, 2020
Revised: April 28, 2021
Accepted: June 7, 2021
Published: June 30, 2021

REFERENCES

Allahyar, A., Vermeulen, C., Bouwman, B.A.M., Krijger, P.H.L., Verstegen, M.J.A.M., Geeven, G., van Kranenburg, M., Pieterse, M., Straver, R., Haarhuis, J.H.I., et al. (2018). Enhancer hubs and loop collisions identified from single-allele topologies. Nat. Genet. *50*, 1151–1160.

Banigan, E.J., van den Berg, A.A., Brandão, H.B., Marko, J.F., and Mirny, L.A. (2020). Chromosome organization by one-sided and two-sided loop extrusion. eLife *9*, e53558.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., and Cavalli, G. (2017). Multiscale 3D genome rewiring during mouse neural development. Cell *171*, 557–572.e24.

Ciosk, R., Shirayama, M., Shevchenko, A., Tanaka, T., Toth, A., Shevchenko, A., and Nasmyth, K. (2000). Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. Mol. Cell *5*, 243–254.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U S A *107*, 21931–21936.

Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.-M. (2019). DNA loop extrusion by human cohesin. Science *366*, 1338–1345.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF binding polarity determines chromatin looping. Mol. Cell *60*, 676–684.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. Cell *149*, 1233–1244.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin domains: the unit of chromosome organization. Mol. Cell *62*, 668–680.

Dowen, J.M., Bilodeau, S., Orlando, D.A., Hübner, M.R., Abraham, B.J., Spector, D.L., and Young, R.A. (2013). Multiple structural maintenance of chromosome complexes at transcriptional regulatory elements. Stem Cell Reports *1*, 371–378.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schujiers, J., Lee, T.I., Zhao, K., and Young, R.A. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. *3*, 99–101.

Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. Nat. Protoc. *7*, 1728–1740.

Fudenberg, G., Abdennur, N., Imakaev, M., Goloborodko, A., and Mirny, L.A. (2017). Emerging evidence of chromosome folding by loop extrusion. Cold Spring Harb. Symp. Quant. Biol. *82*, 45–55.

Ganji, M., Shaltiel, I.A., Bisht, S., Kim, E., Kalichava, A., Haering, C.H., and Dekker, C. (2018). Real-time imaging of DNA loop extrusion by condensin. Science *360*, 102–105.

Gilleland, E. (2013). Two-dimensional kernel smoothing: Using the R package smoothie (No. NCAR/TN-502+STR). https://opensky.ucar.edu/islandora/object/technotes:514.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., et al. (2014). A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. Cell *157*, 369–381.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. Cell *162*, 900–910.

Haarhuis, J.H.I., van der Weide, R.H., Blomen, V.A., Yáñez-Cuna, J.O., Amendola, M., van Ruiten, M.S., Krijger, P.H.L., Teunissen, H., Medema, R.H., van Steensel, B., et al. (2017). The cohesin release factor WAPL restricts chromatin loop extension. Cell *169*, 693–707.e14.

Handoko, L., Xu, H., Li, G., Ngan, C.Y., Chew, E., Schnapp, M., Lee, C.W.H., Ye, C., Ping, J.L.H., Mulawadi, F., et al. (2011). CTCF-mediated functional chromatin interactome in pluripotent cells. Nat. Genet. *43*, 630–638.

Hanssen, L.L.P., Kassouf, M.T., Oudelaar, A.M., Biggs, D., Preece, C., Downes, D.J., Gosden, M., Sharpe, J.A., Sloane-Stanley, J.A., Hughes, J.R., et al. (2017). Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function *in vivo*. Nat. Cell Biol. *19*, 952–961.

Hay, D., Hughes, J.R., Babbs, C., Davies, J.O.J., Graham, B.J., Hanssen, L., Kassouf, M.T., Marieke Oudelaar, A.M., Sharpe, J.A., Suciu, M.C., et al. (2016). Genetic dissection of the α-globin super-enhancer in vivo. Nat. Genet. *48*, 895–903.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hnisz, D., Schuijers, J., Lin, C.Y., Weintraub, A.S., Abraham, B.J., Lee, T.I., Bradner, J.E., and Young, R.A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. Mol. Cell *58*, 362–370.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454–1458.

Holwerda, S.J.B., van de Werken, H.J.G., Ribeiro de Almeida, C., Bergen, I.M., de Bruijn, M.J.W., Verstegen, M.J.A.M., Simonis, M., Splinter, E., Wijchers, P.J., Hendriks, R.W., and de Laat, W. (2013). Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. Nucleic Acids Res. *41*, 6905–6916.

Imakaev, M., Fudenberg, G., McCord, R.P., Naumova, N., Goloborodko, A., Lajoie, B.R., Dekker, J., and Mirny, L.A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. Nat. Methods *9*, 999–1003.

Jonkers, I., Monkhorst, K., Rentmeester, E., Grootegoed, J.A., Grosveld, F., and Gribnau, J. (2008). Xist RNA is confined to the nuclear territory of the silenced X chromosome throughout the cell cycle. Mol. Cell. Biol. *28*, 5583–5594.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Keane, T.M., Goodstadt, L., Danecek, P., White, M.A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. Nature *477*, 289–294.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., Baranasic, D., Arenillas, D.J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W.W., Parcy, F., and Mathelier, A. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res *46*, D260–D266. https://doi.org/10.1093/nar/gkx1126.

Kim, E., Kerssemakers, J., Shaltiel, I.A., Haering, C.H., and Dekker, C. (2020). DNA-loop extruding condensin complexes can traverse one another. Nature *579*, 438–442.

Kolovos, P., Brouwer, R.W.W., Kockx, C.E.M., Lesnussa, M., Kepper, N., Zuin, J., Imam, A.M.A., van de Werken, H.J.G., Wendt, K.S., Knoch, T.A., et al. (2018). Investigation of the spatial structure and interactions of the genome at sub-kilobase-pair resolution using T2C. Nat. Protoc. *13*, 459–477.

Krijger, P.H.L., Geeven, G., Bianchi, V., Hilvering, C.R.E., and de Laat, W. (2020). 4C-seq from beginning to end: a detailed protocol for sample preparation and data analysis. Methods *170*, 17–32.

Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M., et al. (2020). CTCF promotes long-range enhancer-promoter interactions and lineage-specific gene expression in mammalian cells. BioRxiv. https://doi.org/10.1101/2020.03.21.001693.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Lettice, L.A., Daniels, S., Sweeney, E., Venkataraman, S., Devenney, P.S., Gautier, P., Morrison, H., Fantes, J., Hill, R.E., and FitzPatrick, D.R. (2011). Enhancer-adoption as a mechanism of human developmental disease. Hum. Mutat. *32*, 1492–1499.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, Y., Haarhuis, J.H.I., Sedeño Cacciatore, Á., Oldenkamp, R., van Ruiten, M.S., Willems, L., Teunissen, H., Muir, K.W., de Wit, E., Rowland, B.D., and Panne, D. (2020). The structural basis for cohesin-CTCF-anchored loops. Nature *578*, 472–476.

Liu, J., Huang, S., Sun, M., Liu, S., Liu, Y., Wang, W., Zhang, X., Wang, H., and Hua, W. (2012). An improved allele-specific PCR primer design method for SNP marker analysis and its application. Plant Methods *8*, 34.

Liu, N.Q., Maresca, M., van den Brand, T., Braccioli, L., Schijns, M.M.G.A., Teunissen, H., Bruneau, B.G., Nora, E.P., and de Wit, E. (2021). WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. Nat. Genet. *53*, 100–109.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Newkirk, D.A., Chen, Y.-Y., Chien, R., Zeng, W., Biesinger, J., Flowers, E., Kawauchi, S., Santos, R., Calof, A.L., Lander, A.D., et al. (2017). The effect of Nipped-B-like (Nipbl) haploinsufficiency on genome-wide cohesin binding and target gene expression: modeling Cornelia de Lange syndrome. Clin. Epigenetics *9*, 89.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. Cell *169*, 930–944.e22.

Northcott, P.A., Lee, C., Zichner, T., Stütz, A.M., Erkek, S., Kawauchi, D., Shih, D.J.H., Hovestadt, V., Zapatka, M., Sturm, D., et al. (2014). Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. Nature *511*, 428–434.

Pugacheva, E.M., Kubo, N., Loukinov, D., Tajmul, M., Kang, S., Kovalchuk, A.L., Strunnikov, A.V., Zentner, G.E., Ren, B., and Lobanenkov, V.V. (2020). CTCF mediates chromatin looping via N-terminal domain-dependent cohesin retention. Proc. Natl. Acad. Sci. U S A *117*, 2020–2031.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (W1), W160–W165.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Rao, S.S.P., Huang, S.-C., Glenn St Hilaire, B., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin loss eliminates all loop domains. Cell *171*, 305–320.e24.

Ren, G., Jin, W., Cui, K., Rodrigez, J., Hu, G., Zhang, Z., Larson, D.R., and Zhao, K. (2017). CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. Mol. Cell *67*, 1049–1058.e6.

Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.-M., Gibson, T.J., and Zaugg, J.B. (2017). CTCF-mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. Cell Syst. *5*, 628–637.e6.

Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc. Natl. Acad. Sci. U S A *112*, E6456–E6465.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Splinter, E., and de Laat, W. (2011). The complex transcription regulatory landscape of our genome: control in three dimensions. EMBO J. *30*, 4345–4355.

Splinter, E., Heath, H., Kooren, J., Palstra, R.-J., Klous, P., Grosveld, F., Galjart, N., and de Laat, W. (2006). CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. Genes Dev. *20*, 2349–2354.

Splinter, E., de Wit, E., Nora, E.P., Klous, P., van de Werken, H.J.G., Zhu, Y., Kaaij, L.J.T., van Ijcken, W., Gribnau, J., Heard, E., and de Laat, W. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. *25*, 1371–1383.

Sun, F., Chronis, C., Kronenberg, M., Chen, X.-F., Su, T., Lay, F.D., Plath, K., Kurdistani, S.K., and Carey, M.F. (2019). Promoter-enhancer communication occurs primarily within insulated neighborhoods. Mol. Cell *73*, 250–263.e5.

Symmons, O., Pan, L., Remeseiro, S., Aktas, T., Klein, F., Huber, W., and Spitz, F. (2016). The Shh topological domain facilitates the action of remote enhancers by reducing the effects of genomic distances. Dev. Cell *39*, 529–543.

Tolhuis, B., Palstra, R.-J., Splinter, E., Grosveld, F., and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active β-globin locus. Mol. Cell *10*, 1453–1465.

van de Werken, H.J.G., de Vree, P.J.P., Splinter, E., Holwerda, S.J.B., Klous, P., de Wit, E., and de Laat, W. (2012). 4C technology: protocols and data analysis. In Methods in Enzymology, C. Wu and C.D. Allis, eds. (Academic Press), pp. 89–112.

Vermeulen, C., Allahyar, A., Bouwman, B.A.M., Krijger, P.H.L., Verstegen, M.J.A.M., Geeven, G., Valdes-Quezada, C., Renkens, I., Straver, R., Kloosterman, W.P., et al. (2020). Multi-contact 4C: long-molecule sequencing of complex proximity ligation products to uncover local cooperative and competitive chromatin topologies. Nat. Protoc. *15*, 364–397.

Vian, L., Pękowska, A., Rao, S.S.P., Kieffer-Kwon, K.-R., Jung, S., Baranello, L., Huang, S.-C., El Khattabi, L., Dose, M., Pruett, N., et al. (2018). The energetics and physiological impact of cohesin extrusion. Cell *173*, 1165–1178.e20.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. Cell *171*, 1573–1588.e28.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit Polyclonal anti-CTCF	Millipore	Cat# 07-729; RRID:AB_441965
Rabbit Polyclonal anti-RAD21	Abcam	Cat# ab992; RRID:AB_2176601
Chemicals, peptides, and recombinant proteins		
Lipofectamine 2000	Invitrogen	11668019
Critical commercial assays		
NucleoMag PCR Beads	Macherey-Nagel	744100.34
AMPure XP	Beckman Coulter	A63881
SureSelect XT2 Target Enrichment System	Agilent	G9621A
Oxford Nanopore Ligation Sequencing Kit	Oxford Nanopore Technologies	LSK109
Deposited data		
Illumina (4C and capture Hi-C) and Nanopore (MC-4C) sequencing data	This study	GEO: GSE173354
in situ Hi-C, DpnII, mESC, GRCm38	Bonev et al., 2017	4DN: 4DNFIC21MG3U; 4DN: 4DNESDXUWBD9
RAD21 ChIP-seq	Dowen et al., 2013	GEO: GSM824847
CTCF ChIP-seq	Nora et al., 2017	GEO: GSM2609195
RNAPII ChIP-seq	Handoko et al., 2011	GEO: GSM699166
OCT4 ChIP-seq	Whyte et al., 2013	GEO: GSM1082340
SOX2 ChIP-seq	Whyte et al., 2013	GEO: GSM1082341
NANOG ChIP-seq	Whyte et al., 2013	GEO: GSM1082342
MED1 ChIP-seq	Kagey et al., 2010	GEO: GSM560348
H3K27ac ChIP-seq	Creyghton et al., 2010	GEO: GSM594579
NIPBL ChIP-seq	Kagey et al., 2010	GEO: GSM560350
Mouse reference genome assembly GRCm38/mm10	UCSC genome browser / NCBI	GRCm38/mm10
129S1_SvImJ_v1 genome assembly	Mouse Genomes Project / Wellcome Trust Sanger Institute	GCA_001624185.1
CAST_EiJ_v1 genome assembly	Mouse Genomes Project / Wellcome Trust Sanger Institute	GCA_001624445.1
Experimental models: cell lines		
Mouse: 129SvJ/castaneous embryonic stem cells	Jonkers et al., 2008	129/cas
LBΔ (targeted mutant 129 / wild type cas)	This study	10-7D
LBΔ CAS (targeted mutant CAS / wild type 129)	This study	10-B1
LB-inv (targeted mutant 129 / wild type cas)	This study	10-4G
LB-inv CAS (targeted mutant CAS / wild type 129)	This study	10-B8
RBΔ (targeted mutant 129 / wild type cas)	This study	11-2E
SEΔ_1 (targeted mutant 129 / wild type cas)	This study	40-1F
SEΔ_2 (targeted mutant 129 / wild type cas)	This study	40-2H
PΔ (targeted mutant 129 / wild type cas)	This study	41-7A
LBΔ-SEΔ_1 (targeted mutant 129 / wild type cas)	This study	42-8H

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LBΔ-SEΔ_2 (targeted mutant 129 / wild type cas)	This study	42-9G
LBΔ-PΔ (targeted mutant 129 / wild type cas)	This study	43-6B
LBΔ-RBΔtot_1 (targeted mutant 129 / wild type cas)	This study	51-A12
LBΔ-RBΔtot_2 (targeted mutant 129 / wild type cas)	This study	51-F11
LBΔ-mutC_1 (targeted mutant 129 / wild type cas)	This study	52-E3
LBΔ-mutC_2 (targeted mutant 129 / wild type cas)	This study	52-H6
Oligonucleotides		
gRNAs for Cas9 targeting, see Table S1	This study	N/A
ssODN to mutate 5′Slc on 129 allele to "mutC," see Table S1	This study	N/A
ssODN to repair 5′Slc on cas allele to WT sequence, see Table S1	This study	N/A
Primers for allele-specific targeted RNA-seq, see Table S6	This study	N/A
Primers for RT-qPCR, see Table S6	This study	N/A
Primers for allele-specific targeted ChIP-seq, see Table S6	This study	N/A
Primers for ChIP-qPCR, see Table S6	This study	N/A
Biotinylated 120-mer RNA baits for capture Hi-C, see Table S7	This study / Agilent	N/A
Recombinant DNA		
Plasmid: pSpCas9(BB)-2A-Puro (pX459v2)	Ran et al., 2013	Addgene #62988
Software and algorithms		
pipe4C	Krijger et al., 2020	https://github.com/deLaatLab/pipe4C
MC-4C pipeline v1.2	Allahyar et al., 2018	https://github.com/deLaatLab/mc4c_py
BWA aligner v0.7.17 (r1188)	Li and Durbin, 2010	https://github.com/lh3/bwa/
HiC-Pro v.2.11.1	Servant et al., 2015	https://github.com/nservant/HiC-Pro/
Juicebox	Durand et al., 2016	https://github.com/aidenlab/Juicebox/wiki/Download
Bowtie 2 v2.3.5.1	Langmead and Salzberg, 2012	https://github.com/BenLangmead/bowtie2/
SAMtools v1.9	Li et al., 2009	http://www.htslib.org/
DeepTools v.3.3.2	Ramírez et al., 2016	https://deeptools.readthedocs.io/en/develop/
MACS2 v2.2.5	Feng et al., 2012	https://pypi.org/project/MACS2/
FIMO v5.1.1	Grant et al., 2011	https://meme-suite.org/doc/fimo.html
4DN insulation score caller	4D Nucleome Network (4DN)	https://github.com/4dn-dcic/docker-4dn-insulation-scores-and-boundaries-caller
R	The R Foundation	https://www.r-project.org/
Agilent SureDesign	Agilent	https://earray.chem.agilent.com/suredesign/
Excel	Microsoft	https://www.microsoft.com/en-ww/microsoft-365/excel

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Wouter de Laat (w.delaat@hubrecht.eu).

Materials availability

Unique cell lines and plasmids generated in this study will be made available upon request, without restrictions.

Data and code availability

4C and capture Hi-C (Illumina) and MC-4C (Nanopore) datasets generated during this study are available at Gene Expression Omnibus (GEO) under Accession Code GSE173354.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

129SvJ/castaneous female mouse embryonic stem cells (Jonkers et al., 2008; Splinter et al., 2011) were cultured at 37°C and 5% CO_2 on gelatin-coated dishes in the following growth medium: DMEM GlutaMAX (GIBCO), 45% BRL-conditioned medium (11% FBS (Sigma), 1% P/S (GIBCO), 1% NEAA (GIBCO)), 7% FBS (Sigma), 1% P/S (GIBCO), 1% NEAA (GIBCO), 0.1 mM β-mercaptoethanol (Sigma-Aldrich), and 0.2% home-made LIF protein. Medium was changed at least every other day and cells regularly tested negative for *Mycoplasma* infection.

METHOD DETAILS

Genome editing

Cells were transfected in 10-cm dishes using Lipofectamine 2000 (Invitrogen) and 8 μg total DNA per transfection, with 6-12 hours incubation time. We used pSpCas9(BB)-2A-Puro (pX459v2.0, a gift from Feng Zhang (RRID:Addgene_62988)) plasmids with gRNA sequences (Table S1) cloned into its BbsI sites. 24-32 hours after starting transfection, cells were harvested and re-plated at low density in puromycin-containing medium (1.8 μg/mL), which was refreshed after 24 hours and replaced with standard medium after 48 hours. When colonies became clearly visible to the naked eye, they were individually picked and transferred to 96 wells plates for expansion and screening by PCR. Positive clones were expanded and genotyped by PCR and Sanger sequencing.

Cell lines LBΔ-SEΔ, LBΔ-PΔ, LBΔ-RBtotΔ, and LBΔ-mutC were generated by re-targeting the same parent LBΔ clonal cell line. To mutate the 5′Slc CTCF motif (CTGCCGGCAGGAGGCGCTG), we targeted this site in LBΔ cells using a Cas9 gRNA and ssODN repair template (IDT, Belgium) (Table S1). Since the 5′Slc site lies within the coding region of *Slco5a1*, we designed the repair sequence to simultaneously disrupt the CTCF motif and preserve the translation sequence (CTGTCGGCAAGATGCCGAG). When we targeted the LBΔ 129 allele, we did not identify any clones in which the WT sequence was intact on the castaneous allele. Therefore, in cells correctly edited on the 129 allele, we re-targeted the mutated castaneous allele with a gRNA specific to the mutated sequence and restored the WT sequence using another ssODN repair template (Table S1).

SNPs for allele-specific analyses

To differentiate between 129 and castaneous alleles, we identified useful SNPs between the 129S1_SvImJ_v1 and CAST_EiJ_v1genome assemblies from the Mouse Genomes Project at the Wellcome Sanger Institute, as made available within the UCSC Genome Browser. SNPs used for genotyping, allele-specific 4C, targeted RNA-seq, targeted ChIP-seq, and MC-4C were verified in our ES cells by PCR on genomic DNA followed by Sanger sequencing.

Allele-specific targeted RNA-seq

Per sample, one to two million cells were harvested and RNA was isolated using TRIzol reagent (Invitrogen). cDNA was prepared using random primers (Promega) and M-MLV RT (H–) (Promega), using 1 μg RNA input per reaction.

Targeted RNA-seq primers (Table S6) were designed as follows: primer binding sites must not contain any SNPs, primers contained 5′ overhangs that acted as adapters for Illumina sequencing, and an informative SNP had to be present within about 35 nt of the designated "P1" primer, enabling sequencing on both SE and PE platforms. cDNA PCR products, including Illumina overhangs, were designed to be between about 300 and 500 bp in length, to allow for efficient Illumina sequencing. Primer sets either bound non-consecutive exons or spanned exon junctions.

cDNA PCR amplification was performed using Q5 polymerase (NEB) and thermal cycler conditions as specified by the manufacturer (T_a 60°C), with additional reaction conditions summarized in Table S2. PCR products were purified for NGS using Roche HiPure PCR Product purification columns.

To calculate gene expression ratios, we counted the number of times each informative 129- or castaneous-specific polymorphism was detected at a specific position in our NGS reads (see Table S2). We then took the ratio of 129 SNP counts to castaneous SNP

counts (129:cas) for each gene in each cell line and normalized these to average WT 129:cas ratios from replicate WT/WT cells (Table S2; Figure S2A). RT-qPCR recapitulated *Slco5a1* expression results (Figure S2B; Table S3).

Allele-specific 4C-seq

4C template preparation was performed as described in van de Werken et al. (2012), with the following modifications: omission of protease inhibitors in lysis buffer; omission of RNase A at crosslinking removal; replacement of phenol-chloroform and ethanol purifications with magnetic bead purification (Macherey-Nagel NucleoMag PCR Beads), as described in (Krijger et al., 2020). In brief, 8-10 million cells per sample were crosslinked in 2% formaldehyde, after which chromatin was digested with DpnII, followed by dilution and ligation to concatenate DNA fragments that were in close spatial proximity at the time of fixation. Subsequently, crosslinks were removed by Proteinase K treatment, followed by template trimming using Csp6I and re-ligation to produce DNA ''circles'' of a suitable size for PCR amplification, using Roche ELT PCR System with primers specific to the 4C viewpoint.

The allele-specific 4C strategy is similar to that described as PE-4Cseq in Holwerda et al. (2013). Briefly, 4C PCR primers were designed generally in keeping with considerations described in van de Werken et al.(2012), with the following considerations: primer binding sites must not contain any SNPs, and one of the primers must be positioned to allow reading of a SNP in the viewpoint fragment by Illumina sequencing, to facilitate splitting 4C reads by allele. If no SNP was present near the primary cut site, then a SNP following the ''non-read'' primer (P2) was used, in which case paired-end sequencing was required.

4C libraries were sequenced on either Illumina NextSeq or MiniSeq platforms. After demultiplexing, reads were split and allocated as follows: for each of the two possible SNPs, we searched for the SNP and its flanking four nucleotides (text searches for 9-nucleotide sequences). For each SNP, R1 reads were allocated into new FASTQ files that were used for mapping using a recently-improved 4C data processing pipeline (Krijger et al., 2020 and https://github.com/deLaatLab/pipe4C) with the following parameters: normalization to 1 million reads in *cis*, non-blind fragments only, window size 21, top 2 read counts removed. Profile overlays were produced using R (r-project.org).

Allele-specific capture Hi-C

Hi-C template preparation was performed as described in Rao et al. (2014) with modifications as follows. Ten million cells per sample were harvested; two biological replicates were generated per genotype. Cells were fixed with 2% formaldehyde; fixation was quenched by glycine at final concentration of 0.2 M. Five-cutter restriction enzyme XapI (ApoI, ThermoFisher ER1381) (Kolovos et al., 2018) was used for *in situ* digestion (400 U/10 million cells). Digested DNA ends were labeled with biotin-14–dATP (Life Technologies) in a Klenow end-filling reaction. End-repaired, ligated, and reverse-crosslinked DNA was subsequently purified using isopropanol and magnetic beads (Macherey-Nagel NucleoMag PCR Beads). Purified DNA was sheared to 300-500 bp with Covaris, following manufacturer's instructions. After AMPure XP size selection, appropriately-sized ligation fragments marked by biotin were immobilized using MyOne Streptavidin C1 DynaBeads (Invitrogen) and ligated to SureSelect^{XT2} indexed adaptors. The non-amplified, immobilized and indexed Hi-C library was subsequently used as input for target-specific capture using RNA baits.

We designed a custom biotinylated RNA bait library (Table S7) using Agilent SureDesign (https://earray.chem.agilent.com/suredesign/). Biotinylated 120-mer RNA baits were designed to target both ends of ApoI restriction fragments within the targeted region (mm10:chr1:10500000-15500000). 120-mers mapping more than once to the mm10 reference genome were excluded from the final library. The library included unrelated baits for a locus on Chromosome 18 (mm10:chr18:64200000-67200000).

Capture Hi-C of the targeted genomic sequences was carried out with SureSelect^{XT2} Target Enrichment System (Agilent G9621A) with the bait library according to manufacturer's instructions (Agilent Technologies), with the modification of either single-round or double-round capture. For single-round capture, a pre-capture PCR amplification of the immobilized and indexed Hi-C template and a post-capture PCR amplification of the enriched library was carried out with 10 and 12 PCR amplification cycles, respectively. For double-round capture, a pre-capture PCR amplification of the Hi-C template and a post-capture PCR amplification of the single-round enriched library was carried out with 10 and 10 cycles, respectively. The total output of post-capture PCR amplification of the single-round enriched library was further captured the second time with the same bait library followed by final amplification with 12 cycles.

Raw sequence data for biological and technical replicates with the same genotype were combined and subsequently mapped and processed using HiC-Pro v.2.11.1 (Servant et al., 2015) with mm10 as a reference genome. SNPs for the 129S1_SvImJ and CAST_EiJ genomes were extracted from the Mouse Genomes Project (Keane et al., 2011) and N-masked in the reference genome. Reads in which both reads were assigned to the 129 allele or for which one allele was assigned to the 129 allele and for which the other was unassigned were used to generate Juicebox input files (Durand et al., 2016). KR-normalized counts were extracted using Straw (Durand et al., 2016), normalized to 1 million reads, two-dimensional boxcar kernel smoothened using the smoothie package (Gilleland, 2013) and plotted using R (https://www.r-project.org/).

Allele-specific targeted chromatin immunoprecipitation (ChIP)

ChIP was performed using Millipore's protocol with minor modifications. In brief, 10 million cells were fixed with 1% formaldehyde/10% FCS/PBS for 10 minutes at RT, then lysed in cell lysis buffer (10 mM Tris, pH 8.0; 10 mM NaCl; 0.2% NP-40; 10 mM Na-Butyrate; Proteinase inhibitor 1x (Roche)) and nuclei lysis buffer (50 mM Tris, pH 8.0; 10 mM EDTA; 1% SDS; 10 mM Na-Butyrate; Proteinase inhibitor 1x). Isolated chromatin was sonicated to 500-1000 bp, processed on Bioruptor (Diagenode) with 30 s on (High) and 30 s off.

After dilution of SDS to 0.15% using dilution buffer (0.01% SDS; 1.1% Triton X-100; 1.2 mM EDTA; 167 mM NaCl; 16.7 mM Tris-HCl, pH 8.0), chromatin was incubated overnight at 4°C, with 5 μL of CTCF (07-729) and 7 μL Rad21 (ab992) antibodies, respectively. The chromatin-antibody complex was precipitated using protein-G coupled agarose beads (Millipore; 16-201). After washing the chromatin-beads complex, chromatin was eluted from the beads in 1% SDS/0.1 M $NaHCO_3$ for 15 minutes at RT, and purified. ChIP templates were validated for further use by qPCR (Table S5).

After validation qPCR, ChIP templates were used for PCR amplification using primers with Illumina adapters for NGS (Table S6), or a two-step multiplex PCR was used to introduce adaptor sequences (Krijger et al., 2020). In the latter case, the first PCR step (17 cycles) was performed in a multiplex manner by mixing five primer sets per ChIP template. These primers contain 5′ overhangs which correspond to a portion of the sequencing adaptor (primer sequences below). In the second PCR step (20 cycles), overhangs are hybridized by universal primers containing remaining part of the Illumina sequencing adaptor (including indexes). For allele-specific quantification, a validated SNP was included at a maximum distance of 50 bp from the reading primer. All PCRs were performed using Q5 polymerase (NEB) and purified with QIAquick PCR Purification Kit (QIAGEN).

To calculate enrichment ratios, we counted the number of times each informative 129- or castaneous-specific polymorphism was detected at a specific position in our NGS reads. We then took the ratio of 129 SNP counts to castaneous SNP counts (129:cas) for each gene in each cell line and normalized these to average WT 129:cas ratios from replicate WT/WT cells (Table S4).

To demonstrate that our method does not suffer from PCR bias, we amplified sonicated input DNA (no IP) followed by NGS and calculated the ratio of raw 129:cas NGS reads (Table S4; Figure S2D). We then compared these to the same ratios calculated for CTCF and RAD21 immunoprecipitated WT/WT samples. For all 8 targeted sites, input DNA 129:cas ratios were close to 1:1, as expected. This was also true for 7 of the 8 sites in WT/WT ChIPs: closer inspection revealed that the CTCF/cohesin site RB4 contains 2 destabilizing polymorphisms in its CTCF motif, thus likely explaining increased enrichment of CTCF and cohesin at this site on the 129 allele in WT cells (Figure S2E).

Allele-specific MC-4C

MC-4C was performed generally as described in Vermeulen et al. (2020). In brief, 20-40 million cells per sample were harvested and fixed using 2% formaldehyde, after which chromatin was digested using DpnII, followed by dilution and ligation to concatenate DNA fragments that were in close spatial proximity at the time of fixation. Subsequently, crosslinks were removed by Proteinase K treatment, followed by digestion with HindIII and re-ligation. The *Slco5a1* viewpoint (5′Slc) and neighbor fragments were then digested using Cas9 protein and *in vitro* transcribed gRNAs CTATAGCGCCTTTCGATGG, TCTACACCGATTTACACCCT, and ACTAGTGCTAAGTAACCTCG. Primers for the final PCR step were designed to specifically amplify 129 template by forcing the 3′ end of each primer to terminate on a validated SNP and by introducing destabilizing mismatches near the 3′ end of each primer, according to guidelines for allele-specific PCR described in Liu et al. (2012). 129 allele specific amplification was further optimized by performing annealing temperature gradient PCRs in parallel on 129/cas and cas/cas genomic DNA and selecting the optimal annealing temperature which did not produce any product from cas/cas template. For the 5′Slc VP MC-4C PCR we used primers GGACAGCTCCAACCGAAGGCG and GAGTTGAACGCTTCGGCCTTC, where destabilizing mismatches are underlined; PCR conditions were: 98°C 20''; 30 cycles of 98°C 10'', 69°C 20'', 72°C 90''; 72°C 5′.

MinION library preparation and sequencing

MC-4C libraries were size selected at 1.5–8 kb by Pippin HT and then prepared for sequencing using Oxford Nanopore Ligation Sequencing Kit LSK109. Sequencing was performed using R9.4 flow cells (FLO-MIN106), in a single lane of a GridION machine. Of note, we found that loading flow cells with less than the recommended amount (i.e., about 50 ng rather than the recommended 700 ng) produced dramatically higher numbers of informative and unique MC-4C reads. Before commencing each sequencing run, GridION firmware was updated (at time of writing: MinKNOW core 3.6.5, Bream 4.3.16, GUI 3.6.16, Guppy 3.2.10, script configuration 1.0.9). Nanopore sequencing squiggles were converted to bases (i.e., FASTQs) by the GridION interface.

MC-4C data processing

Sequenced reads were processed following the standard MC-4C protocol and MC-4C pipeline v1.2 (Allahyar et al., 2018, https://github.com/deLaatLab/mc4c_py). Fragments were mapped to the reference genome using BWA-SW v0.7.17-r1188 (settings: -b 5 -q 2 -r 1 -z 5 -T 15) (Li and Durbin, 2010). Fragments with mapping quality (MQ) < 20 were considered unmapped. Mapped fragments were extended to the nearest DpnII or HindIII restriction site. Pairs of fragments mapping closer than 20 bp were fused and considered single fragments. PCR duplicates were removed under the premise that far-*cis* or -*trans* fragments can be directly used as unique molecular identifiers (Vermeulen et al., 2020). The region of interest (ROI) used in this study was chr1:12.840-13.185 Mb. Reads with fewer than two fragments within the ROI were discarded as uninformative for multi-way contact association analysis (see below).

MC-4C, region-wide coverage visualization

The ROI was discretized into 200 equally-spaced bins and coverage of each bin was computed by counting the number of fragments in a given dataset that overlaps with the current bin. Each read could only contribute a single fragment to each bin (i.e., two fragments in a read mapping to the same bin were counted as one). Similar to a Hi-C matrix, coverage was represented in a 2-dimensional matrix

where each square represents the number of interactions found between two bins. The resulting matrix was then smoothed using a Gaussian kernel (sigma = 1.0, span = 7 bins). Finally, similar to ICE normalization (Imakaev et al., 2012), the smoothed matrix was normalized to represent coverage probabilities (summation of each row/column = 1).

MC-4C, association analysis (VP-SOI and region-wide)

Similar to Vermeulen et al. (2020), we identified cooperative, random, or mutually-exclusive multi-way contacts between the viewpoint V and two other sites of interest, SOIs X and Y. Briefly, if a preferential three-way contact exists between V, X, and Y, a subset of reads that contain both V and X will frequently cover Y as well. To determine whether Y contact is preferred, undesired, or no preference exists, we compared the frequency of Y in the set of reads that contain both V and X ("positive set of reads") to the frequency of Y in the set of reads that contain V, but not X ("negative set of reads"). The converse analysis, wherein the contact frequency of V and X in presence of Y was compared to the contact frequency of V and X in absence of Y, was also conducted, and in all cases corroborated previously identified (reciprocal) associations. To account for technical and sampling variation, we sub-sampled reads from the negative set to the number of reads in the positive set and repeated this procedure 1,000 times. Finally, the mean and standard deviation of the frequency at which SOI Y was observed in the (sub-sampled) negative set was calculated across the repeated samplings. Using these statistics, z-scores were determined to calculate the significance of preferred or undesired contacts between V, X, and Y. A modest (close to zero) z-score indicates a random contact prevalence between X and Y when V is present. A positive or negative z-score implies a preferred or mutually-exclusive contact between the three elements, respectively.

Reanalysis of ChIP-seq data

Mapping of raw FASTQ files was performed with Bowtie 2 v2.3.5.1 (Langmead and Salzberg, 2012) to the mm10 genome with default settings. The mapped reads with mapping quality score < 15 and PCR duplicates were discarded using SAMtools v1.9 (Li et al., 2009). Bigwig files were generated with DeepTools v.3.3.2 (Ramírez et al., 2016) with bin length of 25 bp, extending reads to 200 bp and RPGC normalization.

CTCF peaks were called using MACS2 v2.2.5 (Feng et al., 2012) at a q-value cutoff of 0.01. CTCF motif and the orientation of each peak was identified using FIMO v5.1.1 (Grant et al., 2011) with motif MA0139.1 (Khan et al., 2018) with -max-stored-scores 50000000.

Insulation score

The insulation score for WT ESCs was calculated using the 4DN insulation-score caller (https://github.com/4dn-dcic/docker-4dn-insulation-scores-and-boundaries-caller) using the default settings (bin size of 1 kb, window size of 100 kb) and 4DNFIC21MG3U.mcool (https://data.4dnucleome.org/files-processed/4DNFIC21MG3U/).

QUANTIFICATION AND STATISTICAL ANALYSIS

For targeted RNA-seq, targeted ChIP-seq, ChIP-qPCR, and RT-qPCR, SD = standard deviation for the number *n* biological and/or technical replicate experiments, as indicated in the Figure legends. SD was calculated using Microsoft *Excel* (command = STDEV).