nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0161-5

Enhancer hubs and loop collisions identified from single-allele topologies

Amin Allahyar[1,2,7], Carlo Vermeulen[3,7], Britta A. M. Bouwman[3], Peter H. L. Krijger[3], Marjon J. A. M. Verstegen[3], Geert Geeven[3], Melissa van Kranenburg[3], Mark Pieterse[3], Roy Straver[1], Judith H. I. Haarhuis[4], Kees Jalink[5], Hans Teunissen[6], Ivo J. Renkens[1], Wigard P. Kloosterman[1], Benjamin D. Rowland[4], Elzo de Wit[6], Jeroen de Ridder[1*] and Wouter de Laat[3*]

Chromatin folding contributes to the regulation of genomic processes such as gene activity. Existing conformation capture methods characterize genome topology through analysis of pairwise chromatin contacts in populations of cells but cannot discern whether individual interactions occur simultaneously or competitively. Here we present multi-contact 4C (MC-4C), which applies Nanopore sequencing to study multi-way DNA conformations of individual alleles. MC-4C distinguishes cooperative from random and competing interactions and identifies previously missed structures in subpopulations of cells. We show that individual elements of the β-globin superenhancer can aggregate into an enhancer hub that can simultaneously accommodate two genes. Neighboring chromatin domain loops can form rosette-like structures through collision of their CTCF-bound anchors, as seen most prominently in cells lacking the cohesin-unloading factor WAPL. Here, massive collision of CTCF-anchored chromatin loops is believed to reflect 'cohesin traffic jams'. Single-allele topology studies thus help us understand the mechanisms underlying genome folding and functioning.

The invention of chromatin conformation capture (3C) technology[1] and derived methods[2] has greatly advanced our knowledge of the principles and regulatory potential of 3D genome folding in vivo. Insights obtained from genome-wide contact maps derived from Hi-C data include the discovery of topologically associated domains (TADs), structurally insulated units of chromosomes of on average a megabase in size[3–5], and of compartments, nuclear environments in which TADs with similar epigenetic signatures spatially cluster[6]. TADs and nuclear compartments are believed to contribute to genome functioning, whereas chromatin loops are thought to influence genome functioning in a more deterministic, direct fashion. Such loops can only be detected when zooming to a much finer scale than whole chromosomes and TADs, either by ultra-deep Hi-C sequencing or by the application of targeted high-resolution approaches such 4C, 5C or capture-C technologies. Chromatin loops include architectural loops, often anchored by bound CTCF proteins, that form structural chromosomal domains[7,8], as well as regulatory chromatin loops that bring distal enhancers in close physical proximity to target gene promoters to control their transcriptional output. Detailed topological studies and genetic evidence have further indicated that individual enhancers can contact and control the expression of multiple genes. Conversely, single genes are often influenced by multiple enhancers[5,9]. Similarly, in population-based assays, individual CTCF sites can be seen contacting multiple other CTCF sites. Based on such observations it has been hypothesized that DNA may fold into spatial chromatin hubs[10,11]. However, current population-based pair-wise contact matrices cannot distinguish clustered interactions from mutually exclusive interactions that independently occur in different cells. To investigate the existence and nature of specific hubs formed between regulatory sequences, CTCF-binding sites and/or genes, targeted high-resolution and high-throughput strategies are needed for detection, analysis and interpretation of multi-way DNA contacts.

Recently, several 3C procedures have been modified for the study of multi-way contacts between selected genes and regulatory sequences, but so far these approaches have been inherently limited in contact complexity, complicating the interpretation of their data[12–15]. At the genome-wide level, recent breakthroughs in the analysis of multi-way contacts have been made. These technologies give insight into the types of genomic sequences that tend to co-occupy nuclear compartments. For example, a new genome-wide approach for multi-contact analysis, called C-Walks (chromosomal walks)[14], gave a glimpse of the nuclear aggregation of genomic loci, indicating that, at the compartment level, cooperative aggregation between dispersed intra- and inter-chromosomal sequences may be rare but may occur, for example, at Polycomb bodies. C-walks, three-way Hi-C contact analysis[15] and genome architecture mapping[16] are all genome-wide methods that do not offer the local coverage necessary to study the functionally most relevant fine-scale topologies formed at individual genes, individual regulatory sequences and individual domain anchors. To enable this analysis and to dissect the spatial interplay between multiple individual regulatory DNA elements and genes, we developed multi-contact 4C sequencing (MC-4C).

[1]Center for Molecular Medicine, University Medical Center, Utrecht University, Utrecht, the Netherlands. [2]Delft Bioinformatics Lab, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, the Netherlands. [3]Oncode Institute, Hubrecht Institute-KNAW and University Medical Center Utrecht, Utrecht, the Netherlands. [4]Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, the Netherlands. [5]Division of Cell Biology, Netherlands Cancer Institute, Amsterdam, the Netherlands. [6]Oncode Institute and Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, the Netherlands. [7]These authors contributed equally: Amin Allahyar, Carlo Vermeulen. *e-mail: j.deridder-4@umcutrecht.nl; w.delaat@hubrecht.eu

Results

MC-4C enables investigation of multi-way DNA conformations. MC-4C is premised on the fact that 3C-based protocols generate aggregates of DNA segments that reside in each other's 3D proximity in the nucleus. These 'DNA hairballs' are created via in situ formaldehyde cross-linking of chromatin, followed by restriction enzyme–mediated DNA fragmentation and proximity-based religation of cross-linked DNA fragments. The resultant DNA concatemers are characteristically sized > 10 kb[17]. Conventional 3C protocols trim these products further to enable efficient analysis of singular ligation junctions only. The MC-4C protocol is designed to keep these concatemers large, enabling the analysis of multi-way contacts for selected genomic sites of interest through third-generation sequencing, such as the Oxford Nanopore Technologies (ONT) MinION. In brief, MC-4C entails the following steps. Like 4C-seq[18] and targeted locus amplification technology[19], MC-4C selectively PCR-amplifies concatemers with primers specific to a fragment of interest (the 'viewpoint'). For this PCR to be sufficiently effective, 3C PCR template in the range of 2–5 kb is made by digesting the large concatemers with a six-cutter restriction enzyme and religation under conditions supporting self-circularization. To reduce prevalent rolling circle amplification and eliminate abundant uninformative undigested products, Cas9-mediated in vitro digestion of the viewpoint fragment (between the inverse PCR primers) and its two neighbor fragments is performed before PCR. After PCR, the product is size-selected (> 1.5 kb) and sequenced on the MinION sequencing platform (Fig. 1a).

An integral component of MC-4C is its elaborate computational analysis strategy (explained in detail in the Methods), which provides the necessary preprocessing of the ONT data and downstream analysis to enable meaningful interpretation of allelic co-occurrence frequencies. To appreciate local multi-way contacts at the level of individual alleles, it is key to filter and select for the informative reads that have two or more contacts within a predefined chromosomal region of interest. Such analysis requires substantial coverage, as reads having less than two local contacts are not informative for our multi-way analysis. To compute reliable statistics, it is also essential to efficiently remove all reads originating from PCR duplicates. For this, we designed a PCR duplicate removal strategy that is guided by cocaptured fragments far outside the region of interest (Supplementary Fig. 1): the chance of independently capturing a given such fragment more than once is extremely small, implying that these sequences can serve as genomically contributed unique molecular identifiers in MC-4C. After this ultraconservative but very reliable PCR filtering strategy, every remaining read represents a unique microtopology derived from an individual allele. MC-4C contact profiles are thus a direct reflection of single allele measurements, which in principle makes them quantitative, albeit limited still by technically inherent variation that may arise from differences in cross-linking, digestion, ligation and mapping ability between fragments.

To explore new biology that may be identified by MC-4C we applied the technique to three different genetic systems. We chose the mouse β-globin and Pcdhα loci, both constituting multiple gene promoters and enhancer and SE elements that act in concert to control defined developmental and cellular expression patterns. We also selected cohesin-looped topological domain boundaries that, upon cohesin stabilization, show extended loops with much more distal anchor sites in population-based Hi-C[20]. We performed a total of 20 MC-4C experiments (27 MinION sequencing runs) to obtain an average of 13,000 individual allelic microtopologies, spanning an average total of 80,000 spatial contacts, per viewpoint (Supplementary Table 1).

Figure 1 summarizes results from a typical MC-4C experiment. Because of PCR, which has a strong bias for small amplicons, and size selection, which we perform to remove the small amplicons before sequencing, the average raw read size is approximately 2 kb (Fig. 1b and Supplementary Fig. 2). Most span three or four spatial contacts, some up to ten (Fig. 1c and Supplementary Fig. 3), with spatial contacts being scored based on ligation events between restriction fragments that are not immediately juxtaposed in the reference genome. To further reduce the effect of PCR-related over- or under-representation of fragments, we divided the region of interest into 200 bins and quantified the relative interaction frequencies per bin. As in all other 3C methods, the great majority of captured sequences (from raw reads) localize to the immediate chromosomal vicinity of the viewpoint (Fig. 1d and Supplementary Fig. 4). The contact profiles derived from sequences directly ligated to the viewpoint (i.e., those that one would analyze in conventional 4C-sequencing) are almost indistinguishable from those created from the indirectly ligated partners (Supplementary Fig. 5). Collectively this indicates that the additional fragments that we capture and analyze by MC-4C are the result of 3D proximity-based ligation events and represent topologically meaningful genomic multi-way contacts made with the viewpoint fragment.

Evidence for an enhancer hub at the β-globin locus. We first studied higher order conformations of the genetically well-characterized mouse β-globin locus. It carries two embryonic globin genes (*Hbb-y* and *Hbb-bh1*) that compete with two downstream adult globin genes (*Hbb-b1* and *Hbb-b2*) for activation by the upstream β-globin superenhancer (SE)[21–23] during development. This SE, also known as the locus control region, is composed of five regulatory elements (hypersensitivity sites (HS) 1–5), of which HS1–HS4 show enhancer activity[24]. Genetic studies in mice further demonstrate that the two developmentally distinct sets of genes compete for activation between sets, but not among members of each set, and that the four enhancer elements of the SE can compensate to a high degree for each other's activity[25,26]. We performed MC-4C experiments in mouse fetal liver, where the adult genes are highly expressed, and in mouse fetal brain, where the β-globin locus is transcriptionally silent. As viewpoints, we included *Hbb-b1*, HS2 and HS5, as well as HS3 exclusively in liver. When all fragments captured by the HS2 experiment are aggregated across all individually analyzed alleles in a so-called overall MC-4C contact profile, we find pronounced and precise interactions with the other SE constituents, as well as with the active gene promoters, specifically in expressing (fetal liver) but not in nonexpressing (fetal brain) primary mouse cells (Fig. 1e). A similarly detailed and tissue-specific topology is appreciable from the overall MC-4C contact profiles that we obtained when using HS5, *Hbb-b1* or HS3 as viewpoints (Fig. 1f and Supplementary Fig. 6). MC-4C therefore accurately recapitulates in a qualitative manner the previously observed conformational features of the β-globin locus[11,27,28] and additionally specifies contacts within the SE with high precision (see also Supplementary Fig. 6). Results were reproducible between biological replicates, even those sequenced on another third-generation sequencing technology (the Pacific Biosciences sequencing platform) (Supplementary Fig. 7a–d). Nevertheless, in our hands the latter platform provided insufficient reads for the generation of robust contact profiles (Supplementary Fig. 7e), which led us to focus on Nanopore sequencing.

To analyze specific multi-way chromatin conformations adopted by the mouse β-globin locus, we selected from each MC-4C dataset the allelic conformations that contain its viewpoint in contact with a second site of interest (SOI). We then quantified and visualized the contact frequencies with the remaining co-occurring sequences. Figure 2a,b shows two examples of such viewpoint–SOI plots (see also Supplementary Fig. 8). The highly localized peaks exactly at the individual enhancer elements of the SE suggest that alleles that fold to have *Hbb-b1* (Fig. 2a) or HS5 (Fig. 2b) in contact with HS2 are likely to also interact with other SE elements. This would be indicative of enhancer hub formation. We tested this with a statistical method

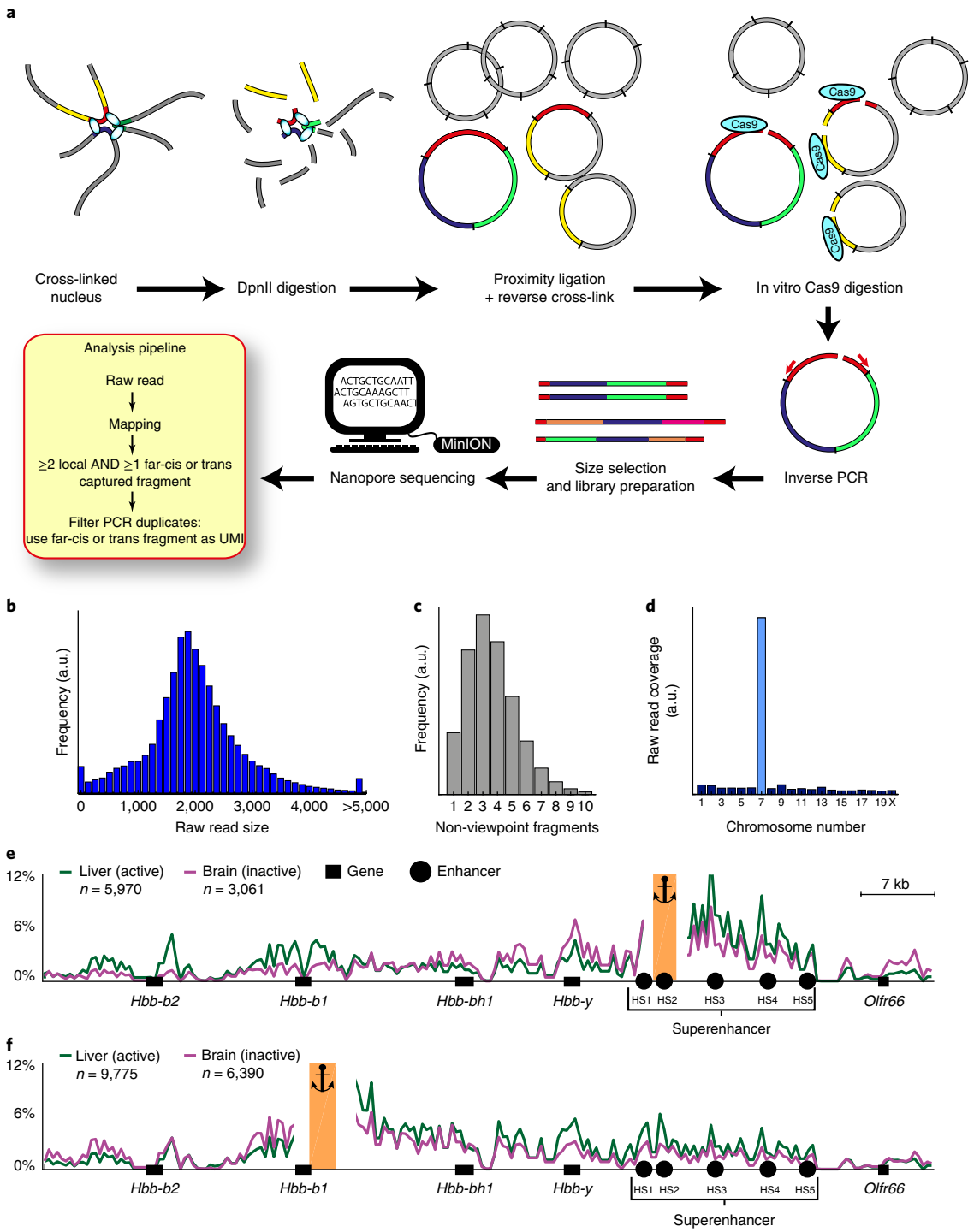

Fig. 1 | Multi-contact 4C technology. a, The MC-4C strategy. **b**–**d**, Statistics of the *Hbb-b1* viewpoint in fetal liver cells. UMI, unique molecular identifier. **b**, MC-4C raw read size distribution after ONT MinION sequencing; a.u., arbitrary units. **c**, The number of MC-4C captured fragments per mapped read, excluding the viewpoint fragment. **d**, Chromosomal distribution of captured and mapped fragments. **e**,**f**, Overall (panallelic) MC-4C contact profile of β-globin HS2 (**e**) and *Hbb-b1* (**f**) in E14.5 fetal liver (green) and brain (purple).

that distinguishes favored from random or disfavored (competitive) multi-way interactions. This method compares through a *z*-score calculation for each sequence its observed three-way co-occurrence frequency with a given viewpoint–SOI combination to its co-occurrence frequency in conformations where the viewpoint is not in contact with the SOI (Fig. 2a,b and Supplementary Fig. 9). By doing so, we analyze whether the chance of being in contact with any third sequence across the region of interest is increased (favored) or decreased (disfavored) when the viewpoint is interacting with a given SOI. Sequences immediately flanking such SOIs are always found to be enriched in this analysis. This is expected as they cannot be spatially separated from the SOI, but we ignore such immediate neighboring sequences here as their favored detection is not reflective of spatial genome organization. Based on our comparative analysis, we find that contacts with the individual elements of the β-globin SE are significantly favored in conformations that already

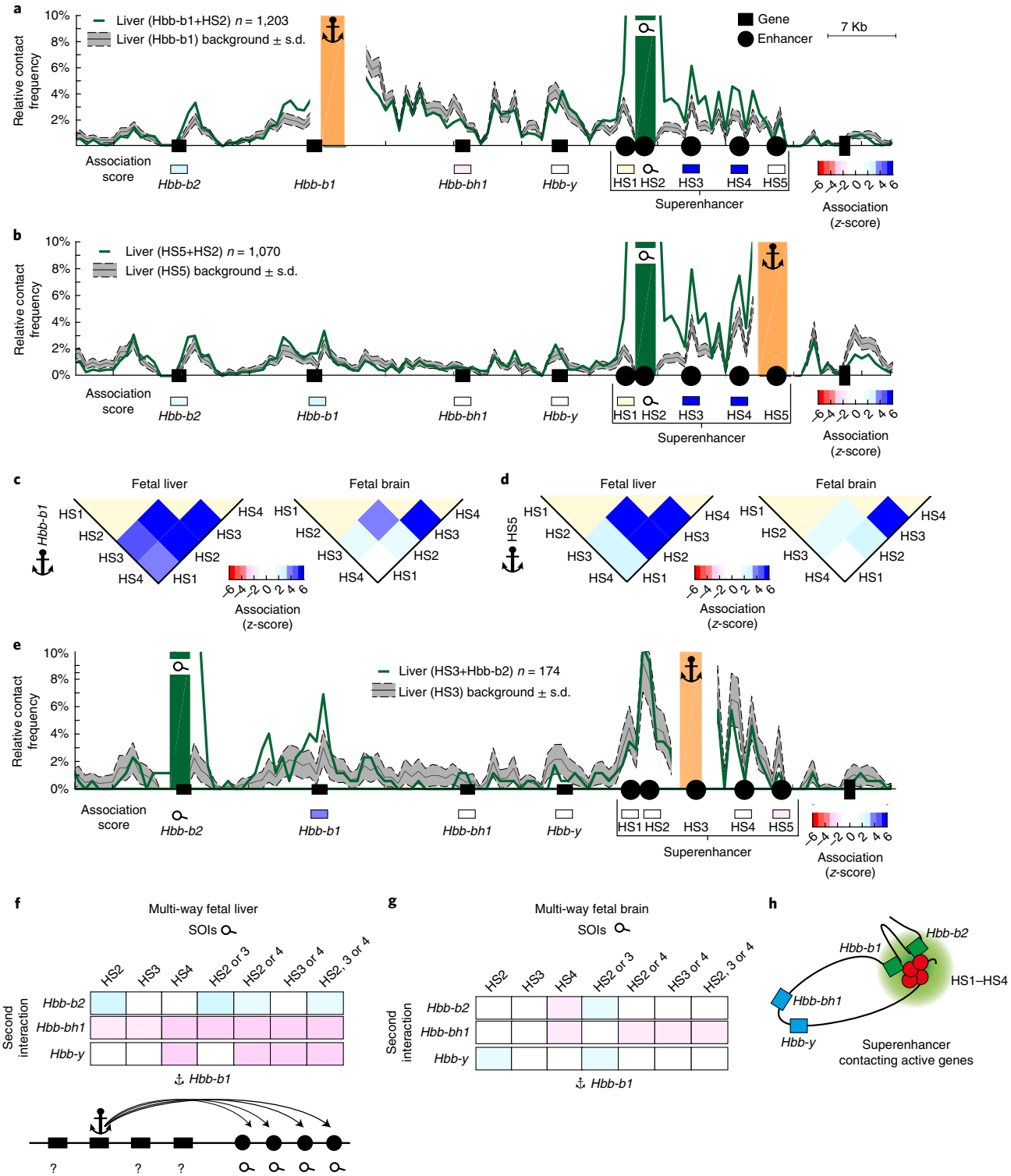

Fig. 2 | A β-globin superenhancer hub that can simultaneously accommodate two genes. a,**b**, Selected microtopologies from fetal liver cells having *Hbb-b1* (**a**) or HS5 (**b**) in contact with HS2 (the number of identified microtopologies is indicated at top left) are specifically enriched for the remainder constituents (HS3 and HS4) of the β-globin SE. Green line shows the observed and gray line the expected (mean ± s.d.) co-occurrence frequency of sites across the locus. *z*-scores (dark blue indicating significant enrichment, dark red indicating significant depletion of a given site at the interrogated microtopology) are shown for SOIs in rectangles below each graph **c**,**d**, Summary of all *z*-scores for all possible pairs of SE elements (HS1 and HS2 are too close to analyze interaction between), when one of them is in contact with the *Hbb-b1* gene (**c**) or HS5 (**d**), in fetal liver (left; β-globin locus active) and fetal brain (right; β-globin locus inactive). Note the preference for co-occurrence between non-neighboring SE elements specifically in fetal liver cells, revealing an active SE hub. **e**, Selected microtopologies from fetal liver cells having HS3 in contact with *Hbb-b2* (the number of microtopologies is indicated in the key). Note that the other active gene, *Hbb-b1*, is preferentially found at these conformations. **f**,**g**, Summary of the clustering behavior (*z*-scores) of the three remainder β-globin genes when *Hbb-b1* is in contact with each of the individual, or combinations of, SE constituents. **h**, Graphic showing the active β-globin SE hub simultaneously contacting the two adult β-globin genes and excluding the two embryonic β-globin genes.

involve one of them. This preferred co-occurrence is appreciable in allelic conformations involving the distal downstream *Hbb-b1* gene, as well as in those involving the upstream HS5 (Fig. 2c). Particularly for the non-neighboring enhancer elements this seems not the result of mere linear proximity, but a consequence of spatial proximity (Fig. 2c and Supplementary Fig. 8). To further rule out the possibility that preferred co-occurrence is a reflection of linear proximity, we repeated the MC-4C experiments on the same locus in nonexpressing tissue (fetal brain). Here no preferred multi-way interactions were observed beyond the directly neighboring constituents (Fig. 2d and Supplementary Fig. 8). This shows that the preferred aggregation of β-globin SE constituents seen in expressing cells is not just the consequence of linear proximity. Preferred clustering of active enhancer elements is found even though these sequences are less cross-linkable when active (formaldehyde-assisted isolation of regulatory elements (FAIRE) identifies enhancers through this principle[29]). We thus conclude that the individual elements of the active β-globin SE can form a higher order enhancer hub.

This SE hub will be visited by the globin genes for their activation. To investigate the number of genes the hub can simultaneously accommodate, we analyzed the likelihood of *Hbb-b2* and the two embryonic globin genes being in contact with the SE when it is interacting with the adult *Hbb-b1* gene (Fig. 2f,g). Despite their linear position between the SE and *Hbb-b1*, the embryonic genes are clearly hindered in contacts with the SE when it is engaged with *Hbb-b1*, particularly in an active tissue (Fig. 2f,g). This suggests that they physically compete with *Hbb-b1* for interactions with the active enhancer hub. For *Hbb-b2*, the other adult globin gene, which is more distal from the SE, we find no indication of physical competition with *Hbb-b1* (Fig. 2e). Its presence is either normally tolerated or even slightly stimulated in topologies having both SE elements and *Hbb-b1* (Fig. 2f). MC-4C therefore provides evidence for two higher order topological phenomena. The first is that the individual elements of a single SE, the active β-globin locus control region, can cooperatively interact (i.e., show statistically increased co-occurrence frequencies) to form a spatial enhancer hub. The second is that this single enhancer hub can physically accommodate two genes simultaneously (Fig. 2h). We find that, in concordance with detailed gene competition studies at this locus[24–26], partnering at the enhancer hub is allowed between developmentally synchronized genes, but not between genes active at different stages of development. These higher order conformational features therefore provide a topological framework that helps interpret genetic observations.

Evidence for an enhancer hub at the Pcdhα locus. Higher order topologies may also help control allelic expression patterns in the mouse protocadherin-α (Pcdhα) gene cluster. Per allele, 1 of 12 alternative promoters (those for *Pcdha1–Pcdha12*) is selected for expression. This ensures that individual neurons express a unique repertoire of membrane-exposed protocadherin molecules, which is essential for axon avoidance[30,31]. Aside from the variable promoters, two constant promoters are active in every neuron (those for *Pcdhac1* and *Pcdhac2*). The activity of nearly all promoters is regulated by two downstream enhancers, HS7 and HS5-1 (only *Pcdhac2* seems not to be influenced by HS5-1)[32,33]. Forward-oriented CTCF binding to all promoters and reverse-oriented CTCF binding to HS5-1 positively contribute to gene expression[34]. Alternative promoter DNA methylation, which prevents CTCF binding, has been proposed to influence allelic promoter choice[35]. We designed viewpoint primers in both enhancers HS5-1 and HS7 and on the promoters of *Pcdha4* and *Pcdha11* and performed MC-4C analysis in mouse E14.5 fetal brain neurons, which express both Pcdhα variants (Supplementary Fig. 10), and in E14.5 fetal liver cells and NIH-3T3 cells, which do not express from any of the Pcdhα promoters. Data from *Pcdha4* and *Pcdha11* and from HS5-1 and HS-7 were pooled owing to the high similarity between overall profiles. All overall contact profiles showed that contacts between the enhancer and each of the promoter regions were perhaps slightly elevated in brain cells, but overall without dramatic differences in locus topology between fetal brain and inactive cells. This suggests that there is no dominant tissue-specific structure conserved in either fetal brain or inactive cells (Fig. 3a,b).

By selectively analyzing the allelic topologies having any of the enhancers in contact with a given alternative promoter in brain cells, we reasoned we could get insight into the specific folding of alleles expressing this particular alternative promoter. As an example, Fig. 3c shows how the other sequences of the locus participate in the microtopologies centered around contacts between the *Pcdha4* or *Pcdha11* promoter, when these are contacting HS7. In neurons, these configurations were specifically enriched for the other enhancer, HS5-1 (39 kb downstream of HS7), as well as for the constitutively active *Pcdhac2* promoter (34 kb upstream of HS7). In liver cells, the corresponding microtopologies did not specifically engage the HS5-1 enhancer, nor any of the genes, as expected if assuming that here these contacts are a reflection of nonfunctional, random collisions. The brain-specific enhancer hub involving cooperative interactions between HS7 and HS5-1 is similarly appreciable when studying other relevant subsets of allelic conformations (Fig. 3d). Additionally, *Pcdhac2* is preferentially found at microtopologies involving interactions between the enhancers and an alternatively transcribed Pcdhα promoter, while *Pcdhac1* is not necessarily evicted from them. The Pcdhα active chromatin hub therefore appears capable of physically accommodating two or more genes at a time. We would have liked to test whether physical competition for enhancer contacts between the *Pcdha1–Pcdha12* promoters may underlie their mutually exclusive allelic expression in neuronal cells. However, the *Pcdha1–Pcdha12* promoters are too close together on the linear chromosome template to observe such mutually exclusive contacts, at least at the current resolution of MC-4C (Supplementary Fig. 11). In summary, as seen for the β-globin SE, the active linearly dispersed individual enhancers HS7 and HS5-1 and the *Pcdhac2* promoter of the Pcdhα locus cooperatively interact to form a tissue-specific active chromatin hub that can simultaneously be contacted by at least one additional gene promoter (*Pcdhac1* or *Pcdha1–Pcdha12*). Notably, our studies on the Pcdhα locus further show that MC-4C can be used to characterize the interaction profiles of rare subpopulations of alleles, identifying topological features that are missed by population-based pairwise contact analysis methods.

WAPL depletion leads to collision of CTCF-anchored loops and to cohesin clustering. As a third model system to study multi-way chromatin interactions, we focused on CTCF- and cohesin-anchored chromatin loops. Cohesin is a ring-shaped protein complex that is necessary to form loops between CTCF-bound domain boundaries[36,37]. The 'loop extrusion' model[38,39] predicts that cohesin forms loops by a process in which the chromatin fiber is pulled through its lumen. The loop is then progressively enlarged until two compatible roadblocks (convergently oriented CTCF-bound sites) are reached, where the loop is stably anchored. Without WAPL, cohesin remains bound to chromatin for longer periods of time, which enables CTCF sites to engage with new CTCF partners over much larger distances, as measured by Hi-C across the population of WAPL-deficient (ΔWAPL) HAP1 (human chronic myeloid leukemia) cells[20]. One possibility is that these additional ultra-long-range interactions are the result of cohesin progressing beyond original CTCF roadblocks to mediate direct pairing between more distal CTCF sites. An alternative explanation would be that distant sites are reeled in through the aggregation of CTCF loop anchors (loop 'collision'), which ultimately brings together distal CTCF sites. Population-based pairwise contact studies cannot distinguish between these two scenarios.

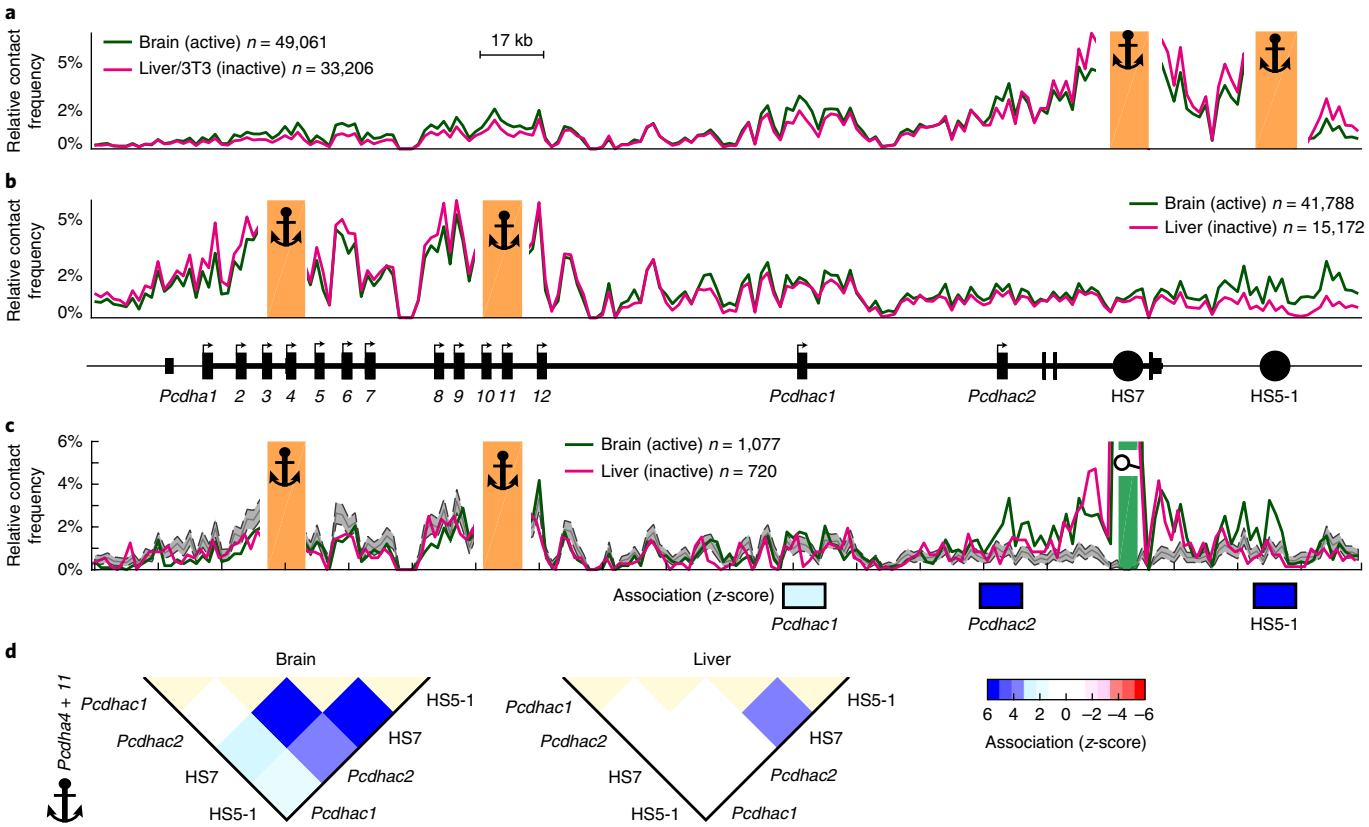

Fig. 3 | MC-4C uncovers Pcdhα hub conformations in tissue-specific subsets of cells. a,**b**, Overall (panallelic) MC-4C contact profiles of the combined HS7 and HS5-1 viewpoints (**a**) and the combined *Pcdha4* and *Pcdha11* viewpoints (**b**) in fetal brain (green; active) and fetal liver or fetal liver and 3T3 cells (purple; inactive). The number of unique reads is indicated in the keys. **c**, Selected microtopologies from fetal brain (green; active) and fetal liver (purple; inactive) having *Pcdha4* or *Pcdha11* (the viewpoints; see anchors) in contact with HS7 (the SOI; see magnifying glass). The gray line and area show the expected (mean ± s.d.) distribution in fetal brain. In fetal brain these presumably are the rare allelic conformations that transcribe either *Pcdha4* or *Pcdha11*. HS5-1 and the pancellularly active *Pcdhac2* gene promoter preferentially cluster at these conformations. The number of identified microtopologies is indicated in the key. **d**, Summary of the clustering behavior (*z*-scores) of the neuron-specifically active Pcdhα elements (*Pcdhac1*, *Pcdhac2*, HS5-1 and HS7) in fetal brain (left, locus active) and fetal liver (right, locus inactive) when any one of these elements is in contact with either *Pcdha4* or *Pcdha11*. Notice the preference for co-occurrence between *Pcdhac2*, HS5-1 and HS7 specifically in the active Pcdhα locus (brain).

MC-4C, which allows quantification of allelic co-occurrence frequencies, does enable disentanglement of these two scenarios.

We selected a region that clearly showed new long-range contacts in ΔWAPL cells based on Hi-C data (Fig. 4a) and applied MC-4C to two CTCF sites that anchor these loops. A comparison between their panallelic contact profiles in wild type (WT) and ΔWAPL cells shows that MC-4C recapitulates the published Hi-C results; it also identifies these long-range contacts specifically in the ΔWAPL cell population (Fig. 4b). If they occur as a result of the skipping of CTCF roadblocks, we would expect a severe depletion of intervening CTCF sites from the allelic microtopologies having these distal CTCF sites together. We find the opposite: intervening CTCF sites show a strong preference to aggregate with these structures, something we observe irrespective of the combination of new long-range contacts we interrogate at this locus (Fig. 4c–d and Supplementary Fig. 12). To exclude the possibility that the effects are locus-specific, we applied MC-4C to another locus showing profound new contacts between distal CTCF sites in ΔWAPL cells. Here as well we find no evidence for mutual exclusivity between CTCF sites that at the cell-population level all seem to interact with each other. Instead, they are again preferentially found clustered at single alleles (Supplementary Figs. 12 and 13). Therefore, rather than—or at least in addition to—the skipping of CTCF roadblocks, our data strongly suggest that WAPL depletion results in loop collision, with distal CTCF sites coming into contact because of progressive aggregation of loop domain anchors. With Hi-C it was also noted that, in the absence of WAPL, contacts between 'illegally' (non-convergently) oriented CTCF sites are more frequently observed[20]. This now seems partially explained as an inevitable result of cluster formation: when three or more CTCF sites form topological aggregates, at least one is in the 'wrong' orientation.

WAPL serves to destabilize, but not to prevent, loop formation, and therefore loop anchor clusters may also exist, albeit less frequently, in WT cells. To investigate this, we selected alleles from WT cells that had the same long-range CTCF contacts interrogated earlier in ΔWAPL cells. Notably, these interactions were too rare in WT cells to stand out in population-based Hi-C and panallelic MC-4C contact profiles (Fig. 4a,b). Strikingly, however, in WT cells these rare allelic conformations also showed a strong enrichment of intervening CTCF-based loop anchors. Quantification of alleles showing simultaneous clustering of three or more distinct CTCF anchors showed an increase from 5.6% to 8.6% (for the downstream viewpoint) and from 6.8% to 10.9% (for the upstream viewpoint) in ΔWAPL as compared to WT cells. We therefore conclude that loop collision and anchor aggregation also occur in WT cells, but less frequently, as a result of the counteracting effect of WAPL (Fig. 4e,f and Supplementary Fig. 13).

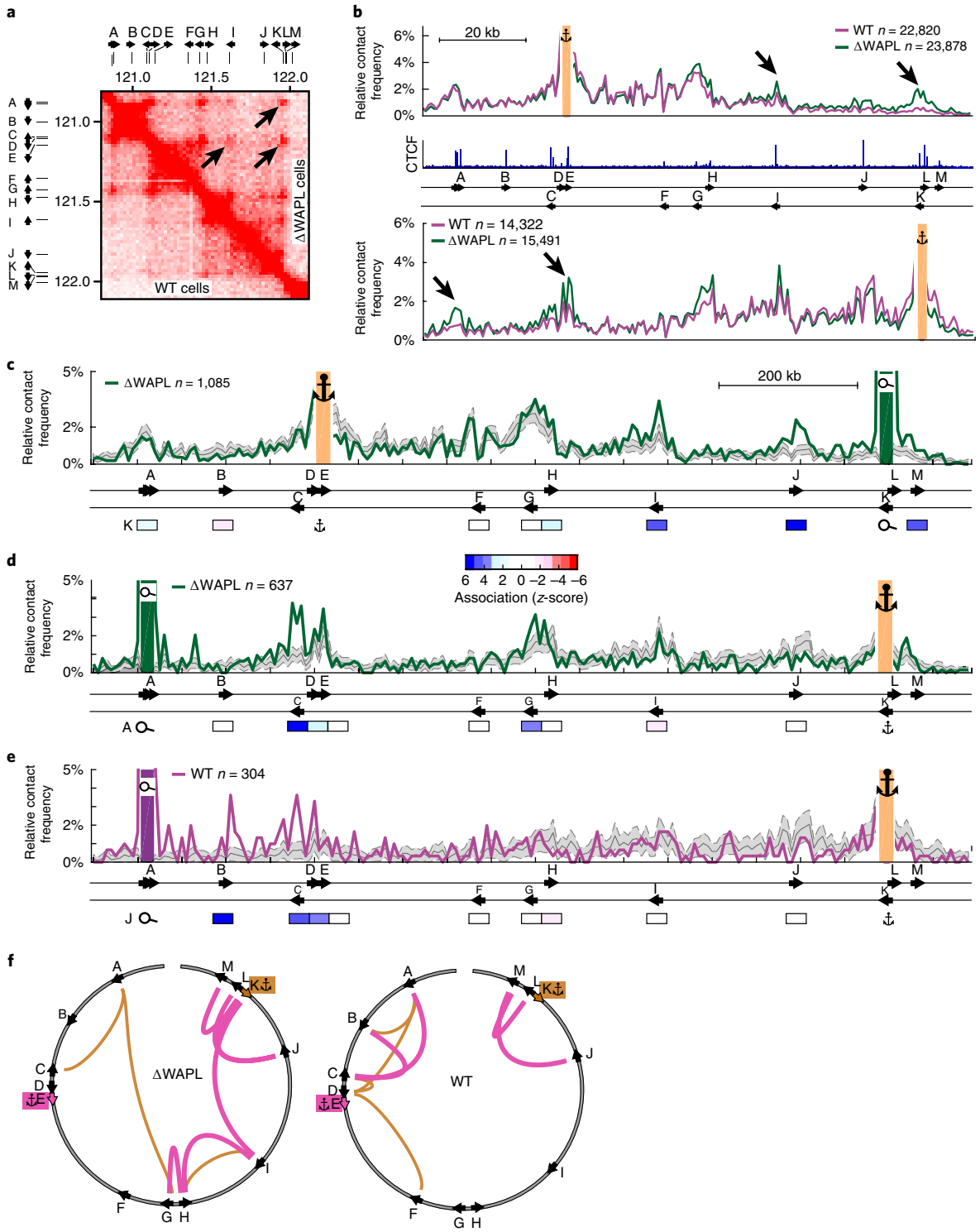

Fig. 4 | Depletion of WAPL stimulates collision of CTCF-anchored domain loops. a, Hi-C contact matrix of a genomic region in wild-type (upper right) and ΔWAPL (lower left) HAP1 cells. Position and orientation of CTCF-binding sites are indicated. Arrows point at new long-range contacts that appear upon WAPL knockout. **b**, Overall MC-4C contact profiles of forward-oriented CTCF site E (top) and reverse-oriented CTCF site K (bottom). The number of unique reads for each experiment is indicated in each plot. ΔWAPL cell CTCF chromatin immunoprecipitation (ChIP)-sequencing profile (from Haarhuis et al.[20]) and CTCF site orientation are shown below. **c**, Microtopologies from ΔWAPL cells having CTCF site E (forward) in contact with CTCF site K (reverse). Gray line and zone indicate negative distribution (mean ± s.d.). *z*-scores are plotted below, showing preferred clustering of CTCF sites I and J. **d**, Selected microtopologies from ΔWAPL cells having CTCF site K (reverse) in contact with CTCF site A (forward). *z*-scores are plotted below, showing preferred clustering of CTCF sites C and G. **e**, Selected microtopologies in WT HAP1 cells having CTCF site K (reverse) in contact with CTCF site A (forward). *z*-scores are plotted below, indicating that the rare allelic conformation wherein K interacts with A co-occurs with interactions with B, C and D, but not with any of the CTCF sites between K and D. **f**, Preferential contacts between CTCF sites. Links are colored with respect to viewpoint and their thickness depicts strength of preferential contacts between CTCF sites.

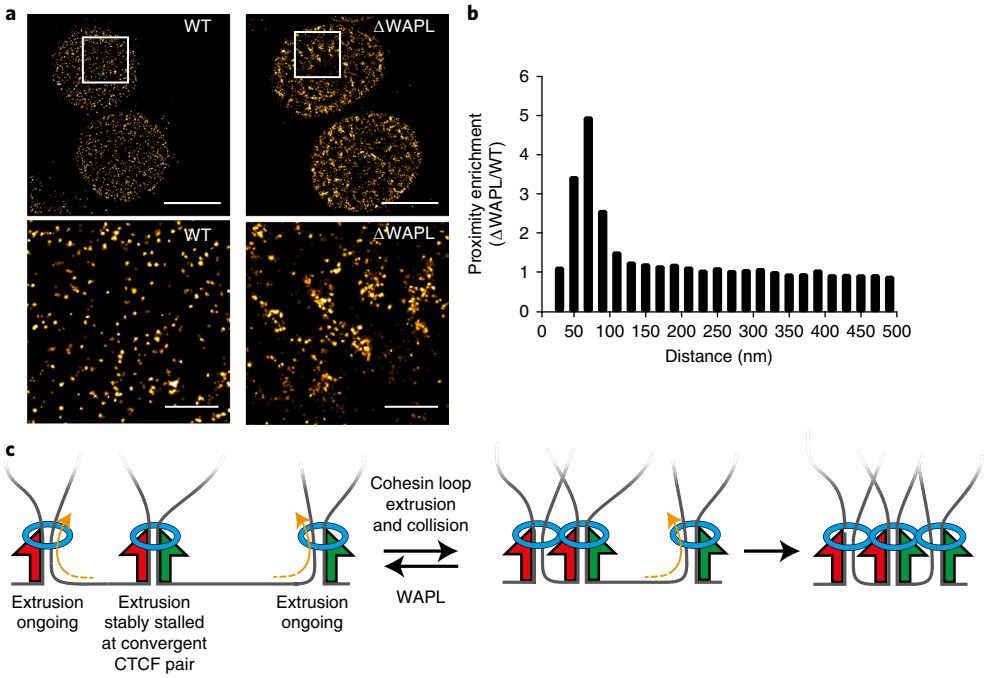

Fig. 5 | Super-resolution microscopy shows cohesin clustering in WAPL-depleted cells. a, Representative example super-resolution images of wild-type and ΔWAPL cells. Scale bars 5 μm (top) and 1 μm (bottom, showing magnifications of boxed regions). The experiment was performed twice, with similar results. **b**, Ratio of the proximity between cohesin particles in wild-type and ΔWAPL cells. For each particle, the distance is measured to all other particles per cell. The graph depicts the proximity enrichment up to a distance of 500 nm. The data shown are the distance measured from 5 cells of each genotype. For dot plots of the distances in individual genotypes, see Supplementary Fig. 13. **c**, Schematic of the proposed traffic jam model explaining the increased incidence of CTCF cluster formation in ΔWAPL cells.

We then searched for an orthogonal methodology that could provide independent evidence for global domain boundary aggregation upon WAPL depletion. For this, we studied the distribution of cohesin in both WT and ΔWAPL cells by means of super-resolution immunofluorescence microscopy. Visual inspection of nuclear images shows a striking reduction of the distance between cohesin molecules in ΔWAPL cells (Fig. 5a). A systematic analysis of their distance distribution patterns confirmed the increased proximity between individual cohesin complexes in these cells (Fig. 5b). Collectively our data strongly suggest that in the absence of WAPL, cohesin-associated domain boundaries massively collide to form rosette-like chromatin structures in interphase nuclei. In light of the loop extrusion model, our findings could be explained by assuming a 'cohesin traffic jam'. Any cohesin ring that is extruding a DNA loop (or sliding over the DNA strands) will eventually be released from DNA by WAPL. If not, it will encounter and presumably be stopped by another cohesin ring that was already immobilized at a CTCF roadblock. Subsequent cohesin rings could then start reeling in other CTCF sites from both directions or as nested loops (loops within larger loops), eventually leading to the spatial aggregation of CTCF-bound loop anchors. Collisions from inside and outside an existing loop would then result in a cohesin traffic jam (Fig. 5c). Although just a theory, loop collisions resulting in a cohesin traffic jam fit well not only with the high frequency of illegal loops seen by Hi-C in ΔWAPL cells but also with the 'vermicelli' cohesin staining patterns observed in ΔWAPL cells[20,40].

Discussion

We present MC-4C, which allows high-resolution analysis of spatial DNA sequence co-occurrence frequencies at the single-allele level. MC-4C contact counts represent relative, not absolute, contact frequencies, as one cannot assume that not being captured (i.e., not being cross-linked, digested, ligated and mapped to the genome) equals not being together. We present a method that, for chosen genomic regions, allows one to statistically distinguish cooperative from random and competitive interactions. In this report we show results directed exclusively toward three-way interactions. Analysis of four-way interactions and beyond poses exponentially increasing demands for the number of analyzed alleles, which is beyond the aims of this study. However, long reads containing more than three fragments are routinely identified, and their content is employed extensively to populate the three-way interaction profiles and to identify PCR duplicates. The data show that, by this method, sequences that directly neighbor each other on the linear chromosome are being scored as obligatorily together in 3D space (cooperative interactions). This is not only as expected (physically connected sequences simply cannot spatially escape each other), but can also be biologically meaningful: it is not without reason that only when transcription factor binding motifs cluster on the linear chromosome can they form functional regulatory motifs. It does emphasize, though, that for correct interpretation of MC-4C results resolution must be high enough to discern spatial clustering as the mere consequence of linear physical proximity from that driven by biological processes. Here we accomplish this by analyzing often more than 10,000 independent allelic conformations per experiment and by comparing allelic co-occurrence frequencies of the same locus in its active versus inactive configuration. The study of higher order chromatin topologies at such high resolution uncovers new biology: individual elements of an SE can aggregate to form an enhancer hub that can accommodate multiple genes simultaneously. Observations such as these highlight the architectural context of SE elements, which combined with their combinatorial deletions will help in understanding their functional hierarchy[21–23]. Similarly, we also find that cohesin drives aggregation of CTCF-bound domain boundaries, which is counteracted by WAPL. Our studies on domain boundary clustering, as well as our work on

Pcdhα, further demonstrate that MC-4C can identify and analyze relevant structures missed by population-based contact methods such as Hi-C or 4C because they are present in only a small percentage of cells. High-resolution multi-way contact analysis methods such as MC-4C promise to uncover how the multitude of regulatory sequences and genes truly coordinate their action in the 3D spatial context of the genome.

For the visualization of co-occurrence frequencies of any site of interest with a given MC-4C viewpoint and the calculation of the significance of such three-way interactions, we refer the reader to the interactive viewer that we made available, together with the data shown in this manuscript (see URLs).

URLs. MC-4C processing pipeline, https://github.com/UMCUGenetics/pymc4c/; MC-4C visualization tool, http://www.multicontactchromatin.nl/; ImageJ macro and corresponding raw images, https://github.com/aallahyar/MC-4C_SRMl; temporal median filter for structured background subtraction, https://github.com/rharkes/Temporal-Median-Background-Subtraction; ImageJ, http://imagej.nih.gov/ij/; Thunderstorm plugin for ImageJ, https://github.com/zitmen/thunderstorm; raw sequencing MC-4C data, https://www.ebi.ac.uk/ena/data/view/PRJEB23327; MC-4C processed data, https://doi.org/10.17632/wbk8hk87r2.1.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0161-5.

Received: 27 February 2018; Accepted: 11 May 2018;
Published online: 9 July 2018

References

1. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
2. Denker, A. & de Laat, W. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* **30**, 1357–1382 (2016).
3. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
4. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
5. Sexton, T. et al. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
6. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
7. Dekker, J. & Mirny, L. The 3D genome as moderator of chromosomal communication. *Cell* **164**, 1110–1121 (2016).
8. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. *Mol. Cell* **62**, 668–680 (2016).
9. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
10. Hanscombe, O. et al. Importance of globin gene order for correct developmental expression. *Genes Dev.* **5**, 1387–1394 (1991).
11. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**, 1453–1465 (2002).
12. Gavrilov, A. A., Chetverina, H. V., Chermnykh, E. S., Razin, S. V. & Chetverin, A. B. Quantitative analysis of genomic element interactions by molecular colony technique. *Nucleic Acids Res.* **42**, e36 (2014).
13. Ay, F. et al. Identifying multi-locus chromatin contacts in human cells using tethered multiple 3C. *BMC Genom.* **16**, 121 (2015).
14. Olivares-Chauvet, P. et al. Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296–300 (2016).
15. Darrow, E. M. et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc. Natl. Acad. Sci.USA* **113**, E4504–E4512 (2016).
16. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).
17. Splinter, E., de Wit, E., van de Werken, H. J., Klous, P. & de Laat, W. Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation. *Methods* **58**, 221–230 (2012).
18. van de Werken, H. J. et al. 4C technology: protocols and data analysis. *Methods Enzymol.* **513**, 89–112 (2012).
19. de Vree, P. J. et al. Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat. Biotechnol.* **32**, 1019–1025 (2014).
20. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707.e14 (2017).
21. Pott, S. & Lieb, J. D. What are super-enhancers? *Nat. Genet.* **47**, 8–12 (2015).
22. Dukler, N., Gulko, B., Huang, Y. F. & Siepel, A. Is a super-enhancer greater than the sum of its parts? *Nat. Genet.* **49**, 2–3 (2016).
23. Huang, J. et al. Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat. Commun.* **9**, 943 (2018).
24. Bender, M. A., Bulger, M., Close, J. & Groudine, M. Beta-globin gene switching and DNase I sensitivity of the endogenous beta-globin locus in mice do not require the locus control region. *Mol. Cell* **5**, 387–393 (2000).
25. Hu, X. et al. Transcriptional interference among the murine beta-like globin genes. *Blood* **109**, 2210–2216 (2007).
26. Cadiz-Rivera, B. et al. The chromatin "landscape" of a murine adult β-globin gene is unaffected by deletion of either the gene promoter or a downstream enhancer. *PLoS One* **9**, e92947 (2014).
27. van de Werken, H. J. et al. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat. Methods* **9**, 969–972 (2012).
28. Davies, J. O. et al. Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80 (2016).
29. Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–885 (2007).
30. Esumi, S. et al. Monoallelic yet combinatorial expression of variable exons of the protocadherin-alpha gene cluster in single neurons. *Nat. Genet.* **37**, 171–176 (2005).
31. Hirayama, T. & Yagi, T. The role and expression of the protocadherin-alpha clusters in the CNS. *Curr. Opin. Neurobiol.* **16**, 336–342 (2006).
32. Kehayova, P., Monahan, K., Chen, W. & Maniatis, T. Regulatory elements required for the activation and repression of the protocadherin-alpha gene cluster. *Proc. Natl. Acad. Sci. USA* **108**, 17195–17200 (2011).
33. Yokota, S. et al. Identification of the cluster control region for the protocadherin-beta genes located beyond the protocadherin-gamma cluster. *J. Biol. Chem.* **286**, 31885–31895 (2011).
34. Guo, Y. et al. CTCF/cohesin-mediated DNA looping is required for protocadherin α promoter choice. *Proc.Natl. Acad. Sci. USA* **109**, 21081–21086 (2012).
35. Toyoda, S. et al. Developmental epigenetic modification regulates stochastic expression of clustered protocadherin genes, generating single neuron diversity. *Neuron* **82**, 94–108 (2014).
36. Sofueva, S. et al. Cohesin-mediated interactions organize chromosomal domain architecture. *EMBO J.* **32**, 3119–3129 (2013).
37. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320.e24 (2017).
38. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
39. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. USA* **112**, E6456–E6465 (2015).
40. Tedeschi, A. et al. Wapl is an essential regulator of chromatin structure and chromosome segregation. *Nature* **501**, 564–568 (2013).

Acknowledgements

We thank the Utrecht Sequencing Facility for providing sequencing data and service, E. Schijlen for help with initial Pacific Biosciences sequencing, and D. Leyton-Puig for help with imaging. We thank N. Geijsen and P. Shang (Hubrecht Institute) for providing Cas9 protein. This work was supported by an NWO VIDI grant (639.072.715) to J.d.R. and an NWO/CW TOP grant (714.012.002) and NWO VICI grant (724.012.003) to W.d.L. and by the NIH Common Fund Program, grant U01CA200147, as a Transformative Collaborative Project Award (TCPA, TCPA-2017-DE-LAAT).

Author contributions

A.A. designed and performed the computational analysis, prepared corresponding plots and wrote the methods and Supplementary Information sections. C.V. and B.A.M.B. designed and performed experiments. C.V. wrote the manuscript and designed figures. P.H.L.K., M.J.A.M.V., M.v.K., M.P. and H.T. performed 'C' methods experiments. R.S. implemented the pipeline in Python. J.H.I.H. generated ΔWAPL cell lines and prepared microscopic slides for super-resolution imaging. K.J. guided acquisition and analyzed super-resolution microscopy data. I.J.R. performed and W.P.K. designed and supervised MinION sequencing experiments. B.D.R. supervised the generation of ΔWAPL cell lines and preparation of microscopic slides for super-resolution imaging. G.G. and E.d.W. helped with computational analysis. E.d.W. performed data analysis on the ΔWAPL

Hi-C data. J.d.R. designed and supervised the computational analyses and pipelines and cowrote the manuscript. W.d.L. conceived and supervised the study and wrote the manuscript.

Competing Interests

C.V., B.A.M.B., P.H.L.K., M.J.A.M.V. and G.G. are shareholders of Cergentis. E.d.W. is cofounder and shareholder of Cergentis. W.d.L. is founder and shareholder of Cergentis. J.d.R. is cofounder and shareholder of Cyclomics.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0161-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.R. or W.L.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Tissue and cell culture. The Animal Welfare Body of the Hubrecht Institute declares that the use of experimental animals is in compliance with the Dutch law on animal experimentation in line with the European Directive (2010/63/EU). The described studies are part of the project entitled "Single-cell quantification of the regulation of gene expression in the first stages of early embryonic development" that was licensed by the Dutch Competent Authority (AVD 801002016728). The described studies and work protocols were subsequently positively reviewed by the Animal Welfare Body of the Hubrecht Institute, listed as HI 17.33.01- AVD80100 2016 728. Mouse embryos were collected from surplus pregnant animals at 14.5 d post conception, and livers and brains were manually dissected. Cells were brought into single-cell suspension in 10% FBS in PBS using a 40-µm strainer. Wild-type HAP1 cells and WAPL knockout HAP1 cells were cultured and harvested as described by Haarhuis et al.[20] and tested negative for mycoplasma contamination.

MC-4C template preparation. MC-4C template was prepared following the regular 4C protocol (described by van de Werken et al.[27] and Splinter et al.[17]), with several adjustments. DpnII (liver and brain) or MboI (HAP1 cells) digestion was performed in a 500-µl volume and the first ligation was performed in a 2-ml volume. After reverse cross-linking, DNA was precipitated using 20 µl NucleoMag P-beads (Macherey-Nagel) and 2 ml 2-propanol, washed twice using 80% ethanol and resuspended in restriction buffer appropriate for the secondary digestion (HindIII in all cases except for the Man1A viewpoint, where SacI was used). After overnight digestion with the second restriction enzyme, the enzyme was heat-inactivated and the template was circularized by diluted ligation (DNA 5 ng/µl). After ligation the DNA was purified using P-beads (10 µl per ml of ligation volume) and two wash steps using 80% ethanol. To remove any remaining P-beads, the template was purified using the Qiagen PCR purification kit.

In vitro Cas9 digestion of MC-4C template. Per viewpoint, three sgRNAs were designed using the ATUM online design tool: one on each flanking fragment and one between the viewpoint primers. gRNA in vitro transcription template was made using a PCR with two partially overlapping primers (as described by Nakayama et al.[41]). In vitro transcription was done using the Megashortscript T7 transcription kit (Invitrogen). RNA was purified with 4×AMPure purification (Agencourt), using DEPC-treated water and avoiding RNase contamination. Purified Cas9 protein was kindly provided by P. Shang and N. Geijsen. Cas9 was preincubated with the appropriate sgRNA. For typical experiments, we preincubated in a 300-µl volume 600 ng of gRNA with 15 pmol of Cas9 protein for 30 min at room temperature. Subsequently, all preincubations were added to 20 µg MC-4C template DNA and incubated for 3–6 h at 37 °C for digestion. After the digestion, Cas9 was inactivated by adding 1/25th volume of 10% SDS and incubating at 70 °C for 5 min. The template was subsequently purified using a 0.6×AMPure purification.

MC-4C viewpoint-specific PCR. Inverse MC-4C primers were designed on DpnII–DpnII fragments, with an approximately 50-bp offset from the restriction sites (where possible), to facilitate viewpoint detection in the analysis pipeline. PCRs were performed in 96 separate 25-µl reactions, using 100 ng of MC-4C template per reaction. PhireII polymerase (Thermo Fisher) was used with the following protocol: 98 °C for 30 s, followed by 31 cycles of 98 °C for 10 s, 57 °C for 20 s and 72 °C for 1.5 min, followed by 72 °C for five min. After PCR amplification, all reactions were pooled and 1 ml was simultaneously purified and size selected using 0.6×AMPure beads.

***Pcdhα* expression analysis.** RNA was isolated from fetal brain cells using TRIzol (Thermo Fisher). cDNA was produced using M-MLV reverse transcriptase (Promega) and random primers (Promega). PCR was performed on cDNA using the primers described by Kaneko et al.[42].

Super-resolution imaging and localization analysis. Cells were grown on poly-L-lysine-covered Ultraclean coverslips (VWR) and treated with 0.1% Triton in PBS for 1 min to extract cohesin that was not bound to DNA. Cells were fixed with 4% PFA in PBS for 7 min and cohesin was stained using a SCC1 antibody (Millipore, cat. no. 05-908). As a secondary antibody, Alexa-647-conjugated goat anti-mouse (Molecular Probes, cat. no. 21235) was used. Specimens were imaged with a Leica SR-GSD 3D microscope using an oxygen-scavenging system (GLOX: 10% glucose plus 0.5 mg/ml glucose oxidase plus 40 µg/ml catalase) supplemented with a reducing agent (cysteamine hydrochloride) at 100 mM. Series of 12,500 raw images were taken at 11 ms exposure time, using 642 nm excitation in high-power mode. Structured background subtraction with a temporal median filter 144 (see URLs for plugin) was performed on the blinking movies using home-built software, and SR images were rendered with the Thunderstorm plugin of ImageJ (see URLs for plugin), using the drift correction option. Images were rendered with 10-nm pixel size, and individual cohesin complexes were identified as small clusters of blinking localizations (20–50 nm diameter). Following automated local thresholding, particle analysis was carried out to discriminate individual cohesin complexes using additional watershedding to separate overlapping particles. Coordinates of the centers of identified particles were exported to Matlab (five super-resolution images each for WT and ΔWAPL) for analysis of distance distributions. The ImageJ macro designed for this analysis along with the raw images is available online (see URLs).

MinION library preparation and sequencing. Pippin HT size selection within a 1.5–8 kb range was performed on PCR products. Subsequently, libraries were prepared using the Oxford nanopore sequencing kits (SQK-NSK007, SQK-LSK108 or SQK-LSK208, depending on the downstream flow cells) and sequenced with the latest available flow cells in the market (namely, R9 (FLO-MIN105), R9.4 (FLO-MIN106) or R9.5 (FLO-MIN107)). Depending on the flow cell, MinION squiggles were converted to bases using either the MinKnow cloud-based software, Metrichor workflow or Albacore basecaller (v1.2.5). FASTQ files are extracted from base-called data using Poretools[43] (version 0.6.0, setting: ---type best).

MC-4C mapping and association analysis: read validity check. To validate fidelity of the sequenced reads, we identified primers as well as their orientations in each read. To this end, Bowtie2[44] v2.2.6 was employed in local alignment mode (settings: -D 20 -R 3 -N 0 -L 15 -i S,1,0.50 --rdg 2,1 --rfg 2,1 --mp 3,2 --ma 2 -a). We allowed 20% mismatches to take into account errors in nanopore sequencing. To improve efficiency of this step, we grouped reads into batches of 10,000 reads and mapped primer sequences to reads within batches in parallel. This step is likely to take about 30 s on average for each batch.

Analyzing primer arrangements in the sequenced reads showed that some reads (~1% on average) are formed by ligation of two or more individual molecules. We therefore implemented a correction procedure in which read-ligation events (i.e., two divergent primers within a read) are identified and reads containing such events are cleaved into two sub-reads. The produced sub-reads are treated as independent reads in downstream analysis. We discarded any reads that contained more than four primers or more than one read-ligation event. These requirements ensured that only those configurations that clearly arise as a result of a read ligation event go through the correction procedure. The produced sub-reads were discarded if their primer configuration did not validate (for example, identification of non-convergent primers on either ends of a read). At this stage, we also discarded any reads (or sub-reads) that were smaller than 500 bp, as they are unlikely to be of sufficient complexity (in terms of the number of fragments) to be informative for multi-contact analysis (see Supplementary Table 1 for corresponding statistics per experiment).

Read splitting. MC-4C reads are expected to be concatemers of multiple distinct fragments and are therefore mapped using an aligner with split-read mapping capabilities (i.e., splitting a single query read and mapping to multiple coordinates). However, as many reads will consist of more than two fragments and splits are expected to occur at known restriction sites in the genome, we pre-split the reads into prospective fragments using the restriction enzyme recognition sequence (Supplementary Fig. 14). This procedure showed improved efficacy in mapping fragments compared to relying only on the split-read mapping capabilities of the aligner (see Supplementary Fig. 15). Due to sequencing errors, extra restriction sites (i.e., observing GATC instead of the correct GAAC) might be erroneously recognized. To consider such cases, the split fragments that map directly adjacent in the reference genome are further fused together in later stages of the pipeline (see below). For the same reason, restriction sites may be missed. In this case, we relied on the split-read capability of the aligner to correctly identify subfragments.

Mapping. To map the partial reads to the reference genome, we used BWA-SW v0.7.16a[45] in SW mode (settings: -b 5 -q 2 -r 1 -T 15). Furthermore, the Z-best heuristic of this aligner is set to 10 (i.e., -z 10). This heuristic increases accuracy of the aligner at the cost of speed. On average, mapping 1 million fragments takes about an hour using a 64-core system running Linux CentOS v7.0. BWA-SW performed best among several tested split-aligners (Supplementary Fig. 16).

Fragment extension and neighbor fusion. Fragments are extended to nearest restriction site (either the four-cutter or six-cutter restriction site) in the reference genome if they are not delimited by restriction sites. Extension is continued to next restriction site in the reference genome if a given fragment is mapped more than 10 bases after an identified restriction site. Any fragments that map closer than 30 bp in the reference genome are fused together and considered to be a single fragment in the rest of analysis. Finally, any fragment with mapping quality below 20 is considered as unmapped.

Duplicate removal. To detect PCR duplicates, we used a conservative approach based on the premise that, in MC-4C, fragments that map far away from the viewpoint are unlikely to be found more than once due to independent ligation events. Therefore, these far-cis or trans fragments can be directly used as unique molecular identifiers. Thus, if these identifier fragments are identified in two reads, those reads are far more likely to be the result of a PCR duplication than of two independent ligation events. A schematic representation of this approach is depicted in Supplementary Fig. 1.

Once a duplicate is found, the read with the fewest local fragments (i.e., fragments that are mapped within the region of interest) is removed. The region of interest is defined as a region around the viewpoint that contains all

expected interacting genetic elements (see Supplementary Table 2 for respective coordinates). Finally, reads that have fewer than two fragments within the region of interest are discarded as they are not informative in multi-way contact analysis.

Association analysis. To identify favored or disfavored multi-way contacts between the viewpoint (V) and two other sites of interest (SOIs), say X and Y, we perform an association analysis as follows. If a favored three-way contact exists between V, X and Y, a subselection of reads that contain both V and X should frequently cover Y as well. To determine whether Y is favored, disfavored or present at background levels, we compare the frequency of Y in the set of reads that contain both V and X (positive selection; Supplementary Fig. 9a) to the frequency of Y in the set of reads that contain V, but not X (negative selection, representing the background interaction pattern of V; Supplementary Fig. 9b). The converse analysis, wherein the contact frequency of V and X in presence of Y is compared to the contact frequency of V and X in absence of Y, is also conducted, and in all cases it corroborated the earlier findings (Supplementary Figs. 8, 11 and 12).

To account for technical and sampling variation that may occur, we subsampled reads from the negative set to the number of reads in the positive set. This procedure was repeated 1,000 times. We implemented a correction for the fact that by definition reads in the positive set have already contributed a fragment to SOI X. Therefore, the positive profile is effectively produced by smaller reads (i.e., one fewer fragment in each read contributes to the interaction profile). Hence, on average, each read in the negative set supplies an extra fragment to the profile compared to reads in the positive set. To compensate for this and ensure that both negative and positive profiles are constructed on the basis of the same distribution in terms of fragments per read, one fragment from each negative read is randomly removed in every subsampling for the negative set.

Finally, the mean and s.d. of the frequency at which SOI Y is observed in the negative set is calculated. Using these statistics, a *z*-score can be determined to estimate significance of the favored or disfavored contacts formed between V, X and Y. While a modest (close to zero) *z*-score indicate a random contact frequency between X and Y when V is present (Supplementary Fig. 9d), a positive or negative *z*-score implies a favored (Supplementary Fig. 9c) or disfavored (Supplementary Fig. 9e) contact between these three elements, respectively.

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

Data availability. All raw sequencing data used in this study are available through European Nucleotide Archive (PRJEB23327). The processed data can be downloaded from Mendeley Data repository (10.17632/wbk8hk87r2.2)

Code availability. The code used in this manuscript is available at GitHub (see URLs).

References

41. Nakayama, T. et al. Cas9-based genome editing in Xenopus tropicalis. *Methods Enzymol* **546**, 355–375 (2014).
42. Kaneko, R. et al. Allelic gene regulation of Pcdh-α and Pcdh-γ clusters involving both monoallelic and biallelic expression in single Purkinje cells. *J. Biol. Chem.* **281**, 30551–30560 (2006).
43. Loman, N. J. & Quinlan, A. R. Poretools: a toolkit for analyzing nanopore sequence data. *Bioinformatics* **30**, 3399–3401 (2014).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
45. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).

nature research

Corresponding author(s): Wouter de Laat & Jeroen de Ridder

Life Sciences Reporting Summary

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

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

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

Experimental design

1. Sample size

   Describe how sample size was determined.

   Sample size in the sense of independent experiments required to observe a certain effect size is not relevant in this context. Instead, we aimed to capture ~10,000 unique molecules per viewpoint as the corresponding general profile demonstrated stable peaks at expected genomic elements.

2. Data exclusions

   Describe any data exclusions.

   No data are excluded in this study.

3. Replication

   Describe the measures taken to verify the reproducibility of the experimental findings.

   We used reciprocal view points to reliably replicate our findings. Moreover, we observe replicates correlate very well.

4. Randomization

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

   Experimental groups were determined by the tissue of origin or the genotype status of the WAPL locus. However, since sample size for each group is 1, no statistical claims are made on the difference between experimental groups, and therefore no randomized assignment to experimental groups has been performed.

5. Blinding

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

   The investigators were not blinded to group allocation. Since sample size for each group is 1, no statistical claims are made on the difference between experimental groups, therefore no blinding has been performed.

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

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

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

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

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	We used Matlab (v2017a) and Python (v2.7.10) to collect, map and analyze MC4C data. All codes are in a private github repository (available upon request), and will be made publicly available upon publication.

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

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party.	no restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	no antibodies were used

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	wildtype and WAPL-KO HAP1 cell lines were obtained from the Rowland lab (NKI, Amsterdam) and described in Haarhuis et al., Cell 2017.
b. Describe the method of cell line authentication used.	Cells were regularly karyotyped, our results recapitulate earlier published results
c. Report whether the cell lines were tested for mycoplasma contamination.	yes, regularly tested
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	no commonly misidentified cell lines were used

Animals and human research participants

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

11. Description of research animals

Provide all relevant details on animals and/or animal-derived materials used in the study.	For some of our experiments we used 'left over' wildtype Black6 mouse E14.5 embryos, isolated from excess pluggings (i.e. pluggings set up, but not used, by others, therefore of pregnant mice that would have been sacrificed irrespectively). Maternal age was between 3 and 6 months.

Policy information about studies involving human research participants

12. Description of human research participants

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