Research

Small chromosomal regions position themselves autonomously according to their chromatin class

Harmen J.G. van de Werken,[1,2] Josien C. Haan,[3] Yana Feodorova,[4,7] Dominika Bijos,[5] An Weuts,[3] Koen Theunis,[3] Sjoerd J.B. Holwerda,[2,8] Wouter Meuleman,[5,9] Ludo Pagie,[5] Katharina Thanisch,[4,10] Parveen Kumar,[3] Heinrich Leonhardt,[4] Peter Marynen,[6] Bas van Steensel,[5] Thierry Voet,[3] Wouter de Laat,[2] Irina Solovei,[4] and Boris Joffe[4,†]

[1]Cancer Computational Biology Center, Erasmus MC Cancer Institute & Department of Urology, Erasmus MC Cancer Institute, Erasmus University Medical Center, 3015 CN, Rotterdam, The Netherlands; [2]Hubrecht Institute-KNAW & University Medical Center Utrecht, 3584 CT Utrecht, The Netherlands; [3]Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Leuven, 3000, Belgium; [4]Department of Biology II, Ludwig Maximilians University Munich, 82152 Planegg-Martinsried, Germany; [5]Division of Gene Regulation, Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands; [6]Human Genome Laboratory, Department of Human Genetics, KU Leuven, Leuven, 3000, Belgium

The spatial arrangement of chromatin is linked to the regulation of nuclear processes. One striking aspect of nuclear organization is the spatial segregation of heterochromatic and euchromatic domains. The mechanisms of this chromatin segregation are still poorly understood. In this work, we investigated the link between the primary genomic sequence and chromatin domains. We analyzed the spatial intranuclear arrangement of a human artificial chromosome (HAC) in a xenospecific mouse background in comparison to an orthologous region of native mouse chromosome. The two orthologous regions include segments that can be assigned to three major chromatin classes according to their gene abundance and repeat repertoire: (1) gene-rich and SINE-rich euchromatin; (2) gene-poor and LINE/LTR-rich heterochromatin; and (3) gene-depleted and satellite DNA-containing constitutive heterochromatin. We show, using fluorescence in situ hybridization (FISH) and 4C-seq technologies, that chromatin segments ranging from 0.6 to 3 Mb cluster with segments of the same chromatin class. As a consequence, the chromatin segments acquire corresponding positions in the nucleus irrespective of their chromosomal context, thereby strongly suggesting that this is their autonomous property. Interactions with the nuclear lamina, although largely retained in the HAC, reveal less autonomy. Taken together, our results suggest that building of a functional nucleus is largely a self-organizing process based on mutual recognition of chromosome segments belonging to the major chromatin classes.

[Supplemental material is available for this article.]

The spatial arrangement of chromatin is linked to most nuclear processes, including transcription regulation, replication, and changes in gene expression during differentiation and development. Based on gene abundance and replication time, chromatin can be roughly subdivided into three major chromatin classes: (1) euchromatin (EC), which is gene-rich and transcriptionally active and replicates in the first half of S-phase; (2) heterochromatin (HC), which is gene-poor and largely transcriptionally silent and replicates in the second half of S-phase; and (3) constitutive heterochromatin (cHC)—gene-depleted silent chromatin replicating at the very end of S-phase. Importantly, the three major chromatin classes have well-defined differential locations in the interphase nucleus.

Chromosomes decondense and establish intra- and interchromosomal contacts during late telophase and early G1. During decondensation, each chromosome fills a certain volume in the nucleus, the so-called chromosome territory (Cremer et al. 2014). Fully decondensed chromosomes are so tightly juxtaposed that they cannot be distinguished after chromatin staining and require in situ hybridization with chromosome paints for their visualization (Cremer et al. 1988; Lichter et al. 1988). However, while chromosomes coalesce with each other, nuclear EC and HC become progressively separated after mitosis. First, neighboring chromosomes join their centromeres (cHC regions) into chromocenters at the nuclear and nucleolar peripheries. Second, HC regions aggregate and form seamless layers below the nuclear envelope, around nucleoli and chromocenters. Finally, decondensed EC regions form a largely continuous neighborhood in the nuclear interior (for reviews, see Joffe et al. 2010; Cremer et al. 2015; Solovei et al. 2016). Since each chromosome possesses HC and EC regions, segregation of HC and EC causes predictably polarized orientation of chromosomes within the nucleus (Croft et al. 1999;

Goetze et al. 2007; Kupper et al. 2007). Separation of the major chromatin classes is evident not only from microscopic (e.g., Bolzer et al. 2005) but also from chromosome conformation capture studies (Lieberman-Aiden et al. 2009; Rao et al. 2014).

The mechanisms of establishing and maintaining the nuclear architecture in mammalian cells remain poorly understood (Bouwman and de Laat 2015; Sexton and Cavalli 2015; Dixon et al. 2016; Pueschel et al. 2016). In particular, we still know little about the mechanisms responsible for the segregation of active and inactive chromatin (Harr et al. 2016). EC and HC regions are interspersed throughout the linear genome (e.g., for review, see Bickmore and van Steensel 2013). Therefore, segregation of the chromatin classes requires that chromosomes are considerably folded and causes their weaving between the EC and HC nuclear compartments (Solovei et al. 2009, 2016; Joffe et al. 2010; Naumova and Dekker 2010). Chromosome folding is assisted by tethering to the nuclear lamina via lamina-associated domains (LADs) comprising up to 40% of the mammalian genome (Guelen et al. 2008; Peric-Hupkes et al. 2010; Meuleman et al. 2013). At least two major tethers of LADs have been identified recently in mammalian cells, the lamin B receptor-dependent and the lamin A/C-dependent tethers. When both tethers are deleted, cHC and HC dissociate from the nuclear envelope and cluster in the nuclear interior, causing a reverse positioning of EC and HC, a phenomenon known as "nuclear inversion" (Solovei et al. 2013). Importantly, the major chromatin classes are also strongly segregated in inverted nuclei. The most extreme case of chromatin segregation is observed in inverted nuclei of nocturnal rod photoreceptors, naturally lacking both peripheral tethers (Solovei et al. 2009; Eberhart et al. 2012, 2013). Absence of peripheral tethering in rod nuclei suggests that lamina scaffolding plays a secondary role in chromatin segregation.

Uncovering the mechanisms that translate the genomic sequence into spatially ordered chromatin is important for our understanding of epigenetic mechanisms ruling the nucleus. We proposed earlier (Krijger and de Laat 2013; Bouwman and de Laat 2015; Solovei et al. 2016) that chromatin regions with homotypic sequences have higher affinity to each other than to regions with heterotypic sequences and hence form more stable contacts. In particular, preferred contacts between long stretches of highly repetitive sequences, such as centromeric or telomeric repeats, can explain ubiquitous clustering of centromeres or telomeres (Weierich et al. 2003). Contacts between shorter and less abundant sequences, such as rDNA gene clusters, can explain fusion of nucleolar organizer regions (Stults et al. 2008; Grob and McStay 2014).

About 40% of the mammalian genome is occupied by interspersed repeats (Jachowicz and Torres-Padilla 2016). The biggest and best-studied families of interspersed repeats are the short interspersed nuclear elements (SINEs), enriched in gene-rich genomic regions, and the retrotransposon-related long interspersed nuclear elements (LINEs) and long terminal repeats (LTRs), both residing preferably within gene-poor regions (see, e.g., Solovei et al. 2016). We hypothesized earlier (Krijger and de Laat 2013; Solovei et al. 2016) that segregation of EC and HC is based on repeat recognition: EC segments recognize each other via enrichment in SINEs, and HC segments recognize each other via LINEs/LTRs and therefore segregate autonomously. The hypothesis, however, was never tested experimentally. Here, we interrogated this question using mouse cells carrying a human artificial chromosome (HAC) that consists of EC and HC segments and a cHC centromeric region.

Results

We sought to directly test whether small chromosomal segments of different chromatin classes are able to correctly segregate in the nucleus and whether they can establish the correct pattern of lamina association. For this purpose, we studied a human chromosome segregating in a xenospecific mouse background. Although cell lines and mice carrying a whole human chromosome have been established (O'Doherty et al. 2005; Wilson et al. 2008), for our analysis we chose a human artificial chromosome to avoid changes in global gene expression in mouse cells, as well as to minimize the input of a human chromosomal context, including control imposed by an entire chromosome territory. For comparison, however, we used orthologous regions of the endogenous mouse and human chromosomes.

Structure of HACs

Two variants of a HAC, a circular (C-HAC) and a linear (L-HAC), were generated as described earlier (Voet et al. 2001, 2003; Weuts et al. 2012). HACs were introduced by microcell-mediated chromosome transfer (Oshimura et al. 2015) into a male mouse ES cell line which was then used for generating chimeras and subsequent offspring that segregated either the C-HAC or the L-HAC as an independent chromosome (Voet et al. 2001; Weuts et al. 2012). To determine the nucleotide sequence of both HACs, we performed paired-end sequencing on DNA extracted from C-HAC– and L-HAC–containing cells, respectively. Both HACs were found to carry a 4.26-Mb region from human Chromosome 1. This region (human HAC orthology region, hHOR) spans from the *MTF2* gene down to the *DPYD* gene (93.6–97.8 Mb) and comprises three structurally different segments: (1) a gene-rich segment encompassing 19 protein coding genes (~2.1 Mb); (2) a gene-poor segment (~0.6 Mb) including the *PTBP2* gene and part of the intron-rich gene *DPYD*; and (3) a gene desert (~1.5 Mb) (Figs. 1, 2A). The three HAC segments also differ in the density of interspersed repeats: the gene-rich segments are enriched in SINEs, whereas the desert and gene-poor segments are enriched in LINEs and LTRs (Supplemental Table S1).

In addition to the above three segments, both HACs include a multi-megabase–sized alpha-satellite region (α-sat) from the centromere region of human Chromosome 20 and a randomly introduced *loxP* cassette. Paired-end mapping analysis revealed that the α-sat sequences were inserted between the 5′ start and the 3′ end of the gene-rich and gene-poor segments in the C-HAC, respectively. In the L-HAC, however, one side of the α-sat region was connected to the same 5′-start of the gene-rich region as in the C-HAC, but the other side of the α-sat region was rearranged to a gene-desert locus and the tip of the 3′ end of the gene-poor segment was lost from the L-HAC (Fig. 1; Supplemental Table S2). The randomly introduced NEO-*loxP*-3′HPRT1 cassette in the C-HAC, used for positive selection of the chromosome following microcell-mediated chromosome transfer and converting the C-HAC into a L-HAC, was found inserted in a different gene-desert segment at 96.5 and 97.0 Mb. Consistent with the model of Cre-*loxP*-mediated telomere seeding and hence linearization of the C-HAC (Weuts et al. 2012), we detected the expected *loxP* crossover between the NEO-*loxP*-3′HPRT1 and the TEL08-cBAp-BLAS-*loxP* cassettes in the L-HAC by read-pair analysis. The newly formed NEO-*loxP*-Blas-telomere cassette in the L-HAC was still connected to the gene-desert locus at 96.5 Mb at the NEO-site and contained telomere sequences at its BLAS-end (Fig. 1). Additionally, in both
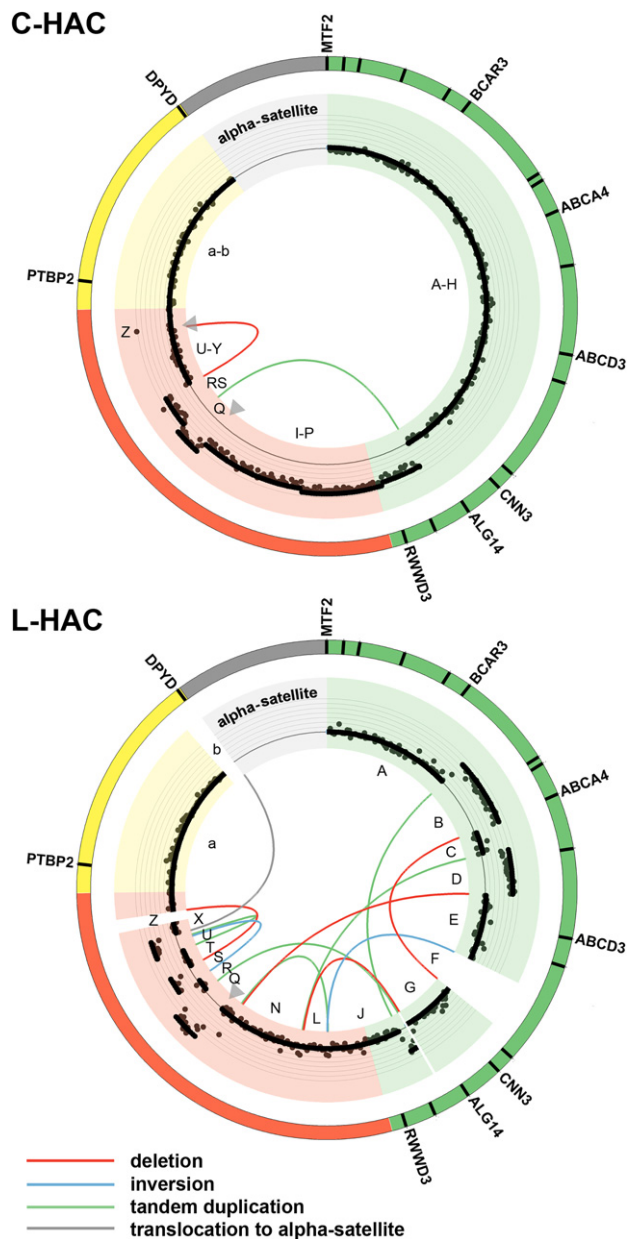
Figure 1. Structural variation of C-HAC and L-HAC. Circos plots show four HAC segments and the structural rearrangements between them. The gene-rich segment is indicated in green; gene-poor, gene desert, and alpha satellite repeat centromeric segments are depicted in yellow, red, and light gray, respectively. The lost sequences are indicated by white gaps. In the C-HAC plot, the NEO-*loxP*-3'HPRT1-construct insertion sites are indicated by gray arrowheads; and in the L-HAC plot, the gray arrowheads point to the location of telomeres. The size of the alpha satellite region is not to scale. Dots in the inner colored circle show the log*R* values per 10-kb bin, solid black lines indicate segments with estimated integer DNA copy number 1, 2, or 3. Chromosome rearrangement signatures are indicated by curved colored lines: deletions, red; tandem duplications, green; inversions, blue; translocations to the centromere region, gray. The letters (A–Z, a-b) indicate names of the blocks specified in Supplemental Table S2. Names of some genes are indicated on the outer colored circle.

HACs, the hHOR region demonstrated multiple rearrangements with concomitant DNA copy number changes—including a large duplication spanning part of the gene-rich and gene-desert seg-

ments in the C-HAC and more complex rearrangements involving the gene-rich and gene-desert segments in the L-HAC, as illustrated in Figures 1 and 2A and Supplemental Table S2. The mouse HAC orthology region (mHOR) on mouse Chromosome 3 has the inverted orientation to the hHOR when comparing both reference genomes. mHOR is syntenic with hHOR in the region from *Dpyd* to *Fnbp1* and also can be subdivided into three segments, gene-rich, gene-poor, and gene desert (Fig. 2B).

Segments within HAC and mHOR faithfully locate in the nuclear zones occupied by the same chromatin classes

For the study of HAC positioning, we have chosen rod photoreceptor cells because the main chromatin classes are clearly segregated in rod nuclei, forming distinct concentric layers. For optical reasons, the arrangement of the three chromatin classes in mouse rods, as in all nocturnal mammals, is inverted in comparison to conventional nuclei of other cell types. Whereas in conventional nuclei, heterochromatin abuts the nuclear and nucleolar peripheries and euchromatin resides in the nuclear interior, in nocturnal rods heterochromatin is concentrated in the interior and euchromatin resides at the periphery (Solovei et al. 2009; Eberhart et al. 2013). In mouse rods, constitutive heterochromatin is packed into a single chromocenter in the very nuclear center. Heterochromatin, depleted of genes and enriched in LINE/LTRs, forms a shell surrounding the chromocenter. Since most of the LINEs belong to the LINE1 (L1) family, we further refer to this shell as the L1-zone. Euchromatin enriched in genes and SINEs forms the outmost peripheral shell surrounding the L1-zone. In mouse, the most abundant SINE family is represented by B1 repeats, and we refer to this shell as the B1-zone (Supplemental Fig. S1).

In mice carrying HACs, some progenitor cells lost the ectopic chromosome and only a proportion of the retinal clones still harbored HACs (Fig. 3A). When visualized by fluorescence in situ hybridization (FISH) using whole human genomic DNA or human Cot1 as a probe, both L-HAC and C-HAC had a rod- or V-shape and were stretched throughout the L1-zone, between the chromocenter and the periphery of rod nuclei (Fig. 3B). To study the localization of the HAC and mHOR segments, we designed cocktail BAC probes encompassing the three noncentromeric segments (Fig. 2A; Supplemental Table S1). For mouse endogenous chromosomes, we additionally designed a probe for the gene-rich segment outside mHOR, upstream of *Dpyd* (Fig. 2B; Supplemental Table S1). All three concentric shells of rod nuclei are easily distinguishable in retina cryosections by differential DAPI staining (Supplemental Fig. S1), and therefore we used confocal image stacks to directly score the positions of HAC and mHOR segments in rods. We found that HAC and HOR segments are faithfully localized in the rod nuclear shells built by chromatin of their own class. In particular, ~90% of the HAC and >80% of mHOR gene-rich segments were found in the B1-zone (Fig. 3C–E). In contrast, 90% of HAC and mHORs gene deserts resided in the L1-zone. Sixty to eighty percent of the gene-poor segments were also associated with the L1-zone (Fig. 3C–E). FISH on retina cryosections did not expose the rearrangements within HACs found by sequencing, most probably due to the low sensitivity of in situ hybridization and the limited resolution of conventional confocal microscopes (Schermelleh et al. 2010; Markaki et al. 2013).

Association of the HAC genic segments with the B1-zone indicated that the HAC genes might be transcriptionally active in mouse cells. Indeed, gene expression analysis by quantitative PCR for genes present in human and mouse orthologous regions
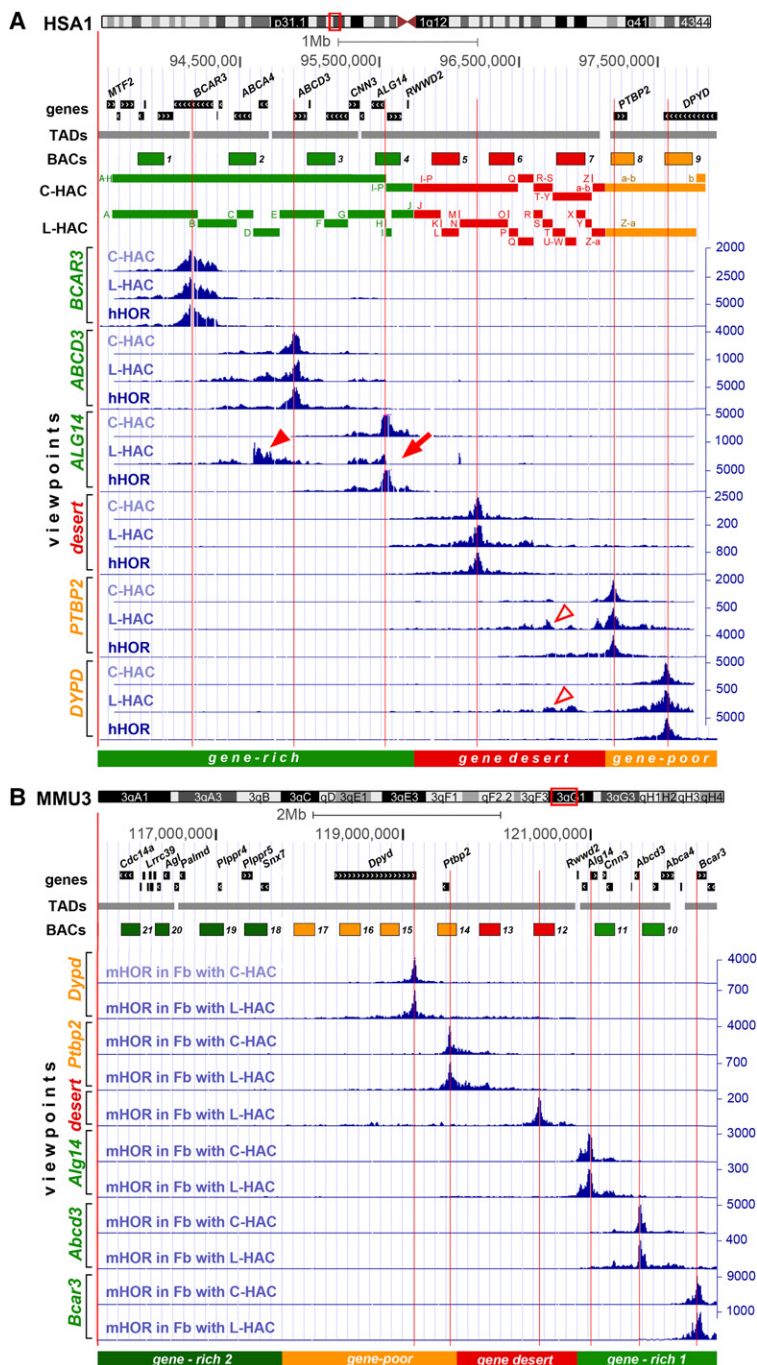
Figure 2. Structure and 4C-seq intrachromosomal contact profiles of the human region (hHOR) within the endogenous Chromosome 1 (HSA1) and HACs (A) and of the mouse endogenous region (mHOR) within Chromosome 3 (MMU3) (B) visualized by the UCSC Genome Browser. The sequence features of the chromosomal loci indicated by red open rectangles in the chromosome ideograms are viewed below the ideograms. Tracks, ordered from top to bottom, are: genes, TADs, BACs used for fluorescence in situ hybridization (FISH) in this study, C- and L-HAC sequence blocks, 4C-seq contact profiles of six viewpoints, division of the regions into segments according to gene abundance. The gene-rich segment is highlighted in light-green; gene-desert and gene-poor segments in red and yellow, respectively. BACs are marked as 1–9 and 10–21 for human and mouse regions, respectively; for the real BAC names, see Supplemental Table S1. An additional flanking gene-rich segment and corresponding BACs for the mouse region are depicted in dark green (B). Note that mHOR has an inverted orientation in comparison with hHOR. Vertical red lines indicate 4C-seq viewpoints. 4C-seq values are shown on the right y-axis in blue. Blocks of rearranged sequences in C-HAC and L-HAC are indicated in A and marked by letters A–Z/a-b as in Figure 1. Arrow and arrowheads point at L-HAC contact profiles regions noticeably altered compared to hHOR as a result of HAC structural rearrangements, such as abrupt depletion of contacts (arrow) and additionally gained contacts (arrowheads).

in A9 cells carrying L-HAC revealed that both HAC and mouse genes were concomitantly expressed (Supplemental Fig. S2). Expression of mouse versus human genes, albeit generally higher for mouse genes, appeared similarly regulated. For instance, highly expressed human *CNN3*, *BCAR3*, and *ALG14* genes corresponded to the highly expressed syntenic mouse genes (Supplemental Fig. S2A,B). However, a direct comparison of human and mouse gene expression was not feasible given the high extent of aneuploidy of both A9 cells and the HAC (Supplemental Fig. S2C–F).

Next, we studied whether human genes are expressed in a tissue-specific manner in mice carrying HACs. To this end, we assessed expression of *ABCA4*, which is active exclusively in photoreceptor cells, and found that human *ABCA4* is expressed along with mouse *Abca4* in retinal but not in liver cells. As expected, the ubiquitously expressed *CNN3/Cnn3* control genes were transcribed in both tissues (Supplemental Fig. S3). The consistently weaker expression levels of *ABCA4* in comparison to *Abca4* probably reflect the difference in their allele number. Thus, genes of human origin are not only transcriptionally active in mouse cells but also manifest a tissue-specific expression, which is in agreement with data on human gene regulation and expression in mouse cells (Tomizuka et al. 1997; Wilson et al. 2008; Kakeda et al. 2011; Kazuki et al. 2013).

In mouse nuclei, centromeres consist of minor satellite repeats and form clusters on the surface of chromocenters (Solovei et al. 2009). In order to define the positions of the HAC centromeres, we cohybridized a probe for a minor mouse satellite with a human pan-centromere alphoid probe. We found that both signals colocalized in ~80%–90% of rods harboring either L-HAC or C-HAC, respectively (Fig. 3G). Thus, human centromeres cluster with mouse centromeres.

Structural domains form independently of their chromosomal context

The strong trend of HAC and mHOR segments to locate within chromatin of their own class in rod cells prompted us to check if the same segregation of chromatin also takes place in cells with conventional nuclear architecture. In conventional nuclei, LINE-rich HC and
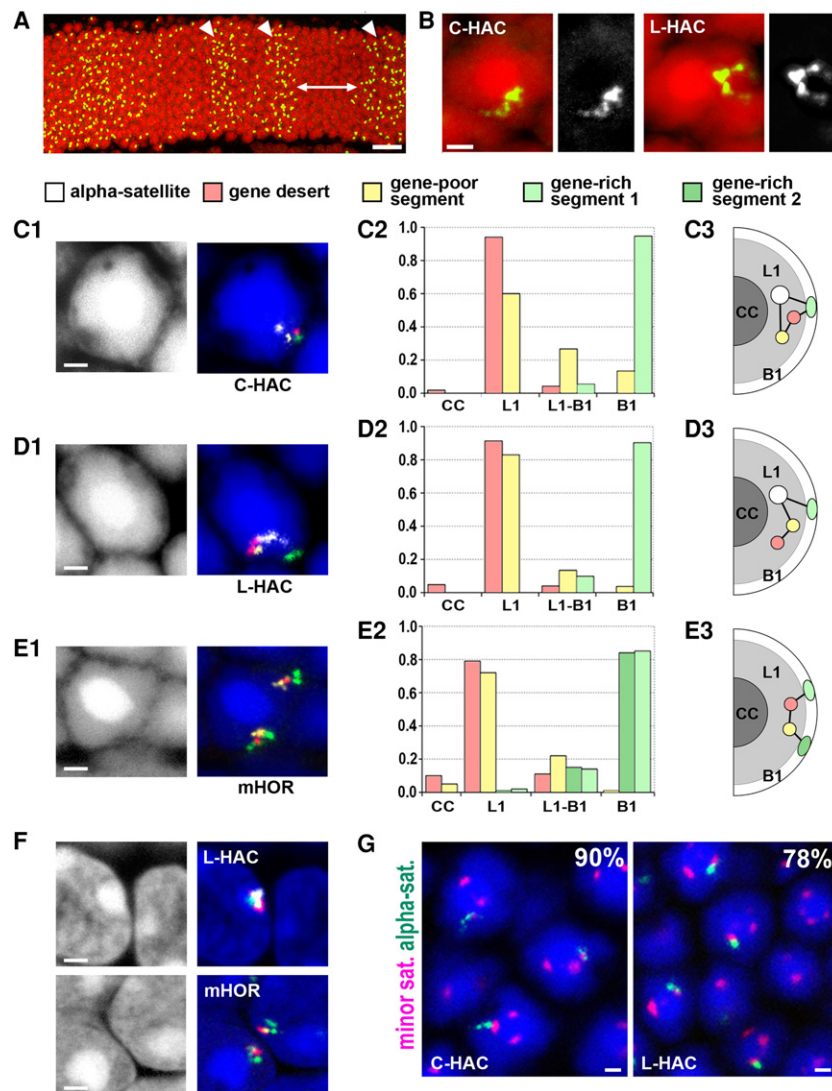
Figure 3. HAC and mHOR segments are faithfully positioned in the rod nuclear shells occupied by the same chromatin class. (*A*) FISH with human genomic DNA as a probe visualizes the HACs in the outer nuclear layer of the mouse retina. Note that some retinal clones are lacking HAC (as the ones marked by the arrow), in others (arrowheads) one HAC per cell is present. (*B*) FISH with human Cot1 as a probe reveals that the HAC is stretched from the central chromocenter to the nuclear periphery (*left*) or has a V-like structure (*right*). DAPI, red; FISH signal, green. Grayscale images of the signals are shown next to the RGB images. (*C*–*E*) Examples of rod nuclei after FISH with a cocktail BAC probe for differential segment labeling (*C1*, *D1*, *E1*), proportion of segments localized in rod nuclear zones (*C2*, *D2*, *E2*), and schematics of typical distributions of the HAC and mHOR segments in rod nuclei (*C3*, *D3*, *E3*). Note that every rod nucleus has one HAC but two alleles of mHOR. The chromocenter (CC), heterochromatic L1-zone (L1), and euchromatic B1-zone (B1) are indicated on the schematics and graphs. The y-axis on the graphs shows the fraction of segment signals from a total number of scored signals for a particular segment. For each segment, between 120 and 250 signals were scored. FISH signals are false-colored as follows: centromere signal, white; gene desert, red; gene-poor segment, yellow; gene-rich segments, green. (*F*) Higher degree of HAC compaction in conventional nuclei of the retinal cells from the inner nuclear layer. (*G*) Examples of HAC and mouse centromere colocalization. Numbers in the corner indicate percentage of rod nuclei with colocalized centromere signals; between 122 and 120 nuclei of rods carrying C-HAC and L-HAC, respectively, were scored. α-satellite, green; minor satellite, magenta. All FISH images are projections of a few optical sections encompassing 1–2 μm. Nuclei in *C*–*G* are counterstained with DAPI (blue). Scale bars: *A*, 20 μm; *B*–*G*, 1 μm.

SINE-rich EC are not packed into well-segregated compartments as in rods but form smaller domains which cannot be reliably discerned by light microscopy (Solovei et al. 2016). Moreover, HACs and orthologous mouse regions in conventional nuclei are much

less stretched in comparison to inverted rod nuclei (Fig. 3F). These two characteristics made a direct FISH signal scoring practically impossible. In a preliminary way, however, segregation of gene-poor and gene-rich segments was indicated by the preferential polar orientation of the studied loci (Supplemental Fig. S4). Indeed, both gene-desert and gene-poor segments favored typical HC environments, such as the nuclear and nucleolar peripheries or the chromocenter surface, whereas gene-dense segments were preferably distanced from these nuclear compartments (Supplemental Fig. S4).

To study the segregation of the segments and their chromatin environment at higher resolution, we applied 4C-seq (chromosome conformation capture combined with high-throughput sequencing) (Splinter et al. 2011; van de Werken et al. 2012a) to cultured fibroblasts derived from mice carrying either the C-HAC or the L-HAC. To this end, we assayed the intra-HAC contact profiles and human-mouse interchromosomal contacts of six sites contained within HACs. These sites included the *BCAR3*, *ABCD3*, and *ALG14* genes in the gene-rich segment, a site inside the gene desert (between the *RWDD3* and *PTBP2* genes), and two genes, *PTBP2* and *DPYD*, in the gene-poor part of the HAC (Fig. 2A). For comparison, we studied the contacts of the same target sequences in their endogenous chromosomal context in human lung fibroblasts (Fig. 2A) and also applied 4C-seq to corresponding sites in mHOR in the mouse fibroblasts carrying HACs (Fig. 2B).

First, we focused on the human orthologous region contacts formed by the target sequences in their natural chromosomal context. We noticed that the gene-dense segment formed a separate structural domain with lots of intra-domain contacts and with very few contacts with the nearby gene desert and gene-poor segment carrying *DPYD* (see Supplemental Fig. S5). Likewise, sequences in the gene-poor segment and in the gene desert, as judged from three independent viewpoints, frequently contacted each other but made very few contacts with the gene-rich segment (Figs. 2A, 4). A similar compartmentalization was seen for mHOR segments in the mouse endogenous chromosomal context in both cell lines carrying either C-HAC or L-HAC (Fig. 2B; Supplemental Fig. S6). This showed that active and inactive parts of the HOR spatially segregate in their human and mouse endogenous chromosomal environment.
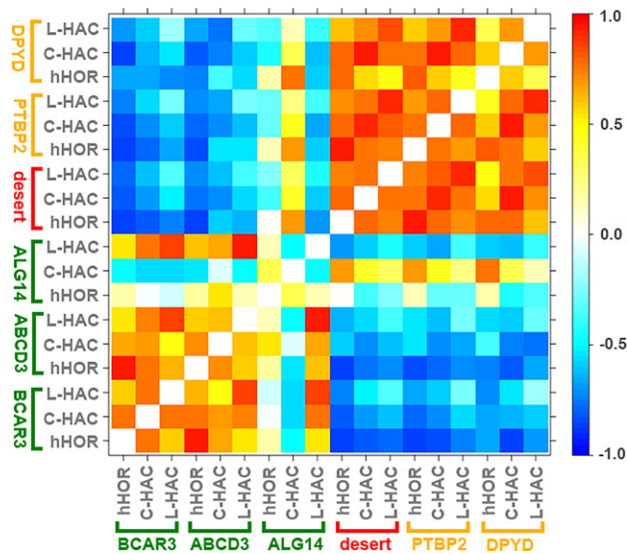
Figure 4. Heat map of Spearman's rank correlation coefficients of the 4C-seq profiles of hHOR, C-HAC, and L-HAC. The corresponding 4C-seq profiles are shown in Figure 2A. 4C-seq data within 200 kb from each viewpoint were excluded. The colors range from anti-correlating (−1.0) blue to correlating red (1.0).

Compartmentalization was seen in both linear and circular HACs (Figs. 2A, 4). The topology of chromatin in a circular configuration is interesting to consider. A previous study showed that the similarly sized (~4 Mb) circular genome of *Caulobacter crescentus* adopts an ellipsoidal structure with periodically arranged arms that zipper up through contacts across the arms. This topology was the consequence of a defined site being anchored in each cell (Umbarger et al. 2011). We found no such configuration and score very few contacts between sequences on opposite sides of the circle, which may indicate that C-HAC is not site-specifically attached to any particular structure in the mouse nucleus. Instead, we found again the strict spatial separation between the gene-dense segment and the neighboring gene desert, as was found for L-HAC and also for mHOR and hHOR in the endogenous chromosomal context (Figs. 2, 4; Supplemental Fig. S6).

Collectively, in agreement with our microscopic observations, the 4C-seq analysis of intra-HOR contacts formed by a genomic region in and outside its chromosomal context further

demonstrates that structural domains form autonomously to physically separate active from inactive chromatin. In addition, although our 4C-seq experiments further expose the rearranged nature of HACs, particularly of the L-HAC, with deletions and inversions being discernible from the contact profiles (arrows and arrowheads on Fig. 2A), they also emphasize a high robustness of contact repertoire of the viewpoints seeded in both gene-rich and gene-poor environments (Fig. 4).

Bipolar HAC orientation places each structural domain in distinct nuclear compartments

We then wondered if the two HACs preferentially positioned themselves near certain regions of the mouse genome in the nuclear space of their host cells. Visual inspection of contacts made across the mouse chromosomes revealed no interaction hotspots and none with the homologous region on Chromosome 3 and vice versa (Supplemental Fig. S7). The number of interacting regions in *trans* already seems to identify the type of chromosomal site that is interrogated, since this number noticeably declines from gene-rich to gene-poor and gene-desert segments (Supplemental Fig. S7). We defined regions of increased contacts and asked whether they had distinguishable sequence properties, perhaps in common with regions contacted by the endogenous mHOR sites. For this, we focused first on the contacts made by *BCAR3*, located in the gene-rich segment of the hHOR, and *DPYD*, a site within the gene-poor segment. The endogenous human genes made interchromosomal contacts with similarly typed genomic regions. Thus, *BCAR3* predominantly contacted GC-rich, SINE-rich, and LINE/LTR-poor regions, and vice versa, *DPYD* was mostly surrounded by GC-poor, SINE-poor, and LINE/LTR-rich regions (Fig. 5, black dots). A similar interchromosomal contact repertoire was found for the endogenous mouse genes *Bcar3* and *Dpyd* in mHOR (Fig. 5, green dots). These observations were in line with the spatial separation of active A and inactive B compartments, which are generally characterized by enrichment in SINEs (hence, in GC) or LINEs (hence, in AT), respectively (Simonis et al. 2006; Lieberman-Aiden et al. 2009; Solovei et al. 2016). Importantly, the mouse interchromosomal regions contacted by the human homologous sequences on the HACs showed highly similar characteristics: *BCAR3* was found mostly in active mouse chromosomal regions, whereas *DPYD* positioned itself predominantly in the inactive compartments (Fig. 5, red and orange dots). We did not find specific *trans*-interactions of the HAC loci with orthologous mouse
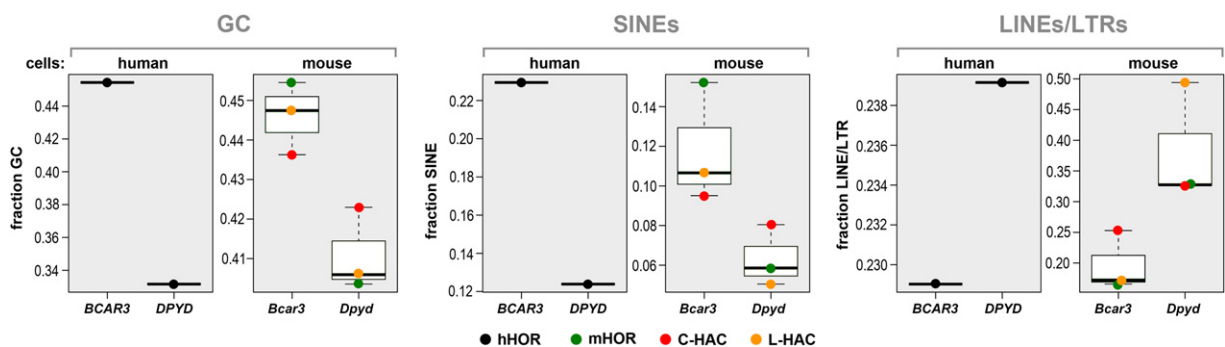

Figure 5. Box plots showing the relative frequency of GC bases, SINE, and LINE/LTR elements of the *trans* interacting 4C-seq regions in human and mouse cells. Dots represent the characteristics of the 4C-seq *trans* interacting regions using different viewpoints on different chromosomes: hHOR, black; mHOR, green; L-HAC, orange; C-HAC, red. Median values are shown as horizontal black lines. *BCAR3/Bcar3* genes represent the gene-rich segment and *DPYD/Dpyd* genes represent the gene-poor segment.

genome loci but revealed HAC bipolar orientation, with its active and inactive segments preferably contacting active and inactive mouse genome regions, respectively (Supplemental Fig. S8). The fact that each gene showed such pronounced and different contact preferences strongly suggested that this is not a technical artifact but a reflection of nuclear organization.

Thus, distinct structural domains located on small artificial chromosomes can independently position themselves in nuclear compartments occupied by similarly typed chromatin domains of the host genome. The fact that both the L-HAC and the C-HAC find bipolar nuclear orientation of active and inactive chromatin further confirms the autonomous capacity of genomic regions to orient themselves in the nucleus. Rather than searching for certain genomic sequences or chromosomal segments, it seems that they search for their preferred genomic environment, characterized by certain repeat repertoire and chromatin composition, which can be found on any given chromosome.

NL association of L-HAC is partially preserved in cultured mouse cells

Finally, we investigated the pattern of interactions of the HACs with the nuclear lamina (NL) using the DamID technique (van Steensel et al. 2001). We focused on HAC–NL interactions in mouse fibroblasts, because NL interaction profiles of normal chromosomes in human and mouse fibroblasts were already available for comparison (Guelen et al. 2008; Peric-Hupkes et al. 2010). We conducted duplicate DamID experiments by expressing Dam-Lamin B1 (Peric-Hupkes et al. 2010) in mouse A9 cells carrying the L-HAC. Cells expressing unfused Dam served as a normalization control. We hybridized the adenine-methylated DNA to genomic tiling arrays querying either human Chromosome 1 or the entire mouse genome.

The DamID profile of human Chromosome 1 in the A9 cells shows a distinctive pattern of high and low signals (Fig. 6A). This pattern is strictly confined to the HAC region, demonstrating that the DamID signal is specifically derived from this HAC. The magnitude of the DamID signal on the HAC is, however, lower than on the homologous mouse region in the same cells (Fig. 6B) and also lower than previously observed for Chromosome 1 in human fibroblasts (Guelen et al. 2008) (cf. Fig. 6, A and B). This suggests that NL interactions of the HAC are generally weaker in mouse cells than in a native chromosome context. In agreement with this, we found by FISH that in A9 cells ~40% of the desert segments in L-HAC associate with chromocenters rather than with the nuclear envelope (Supplemental Fig. S9). Possibly, the chromocenters and the NL compete for HAC interactions.

A region spanning from ~95.6 to 96.7 Mb (genomic coordinates) on the linear HAC exhibits a consistently positive DamID signal. This region overlaps with a strong LAD that is present in both mouse A9 and human fibroblasts (Fig. 6). This suggests that, in the context of the HAC, this region has retained some of its ability to associate with the NL. A few other smaller regions on the HAC also show positive DamID signals, but the homologous sequences
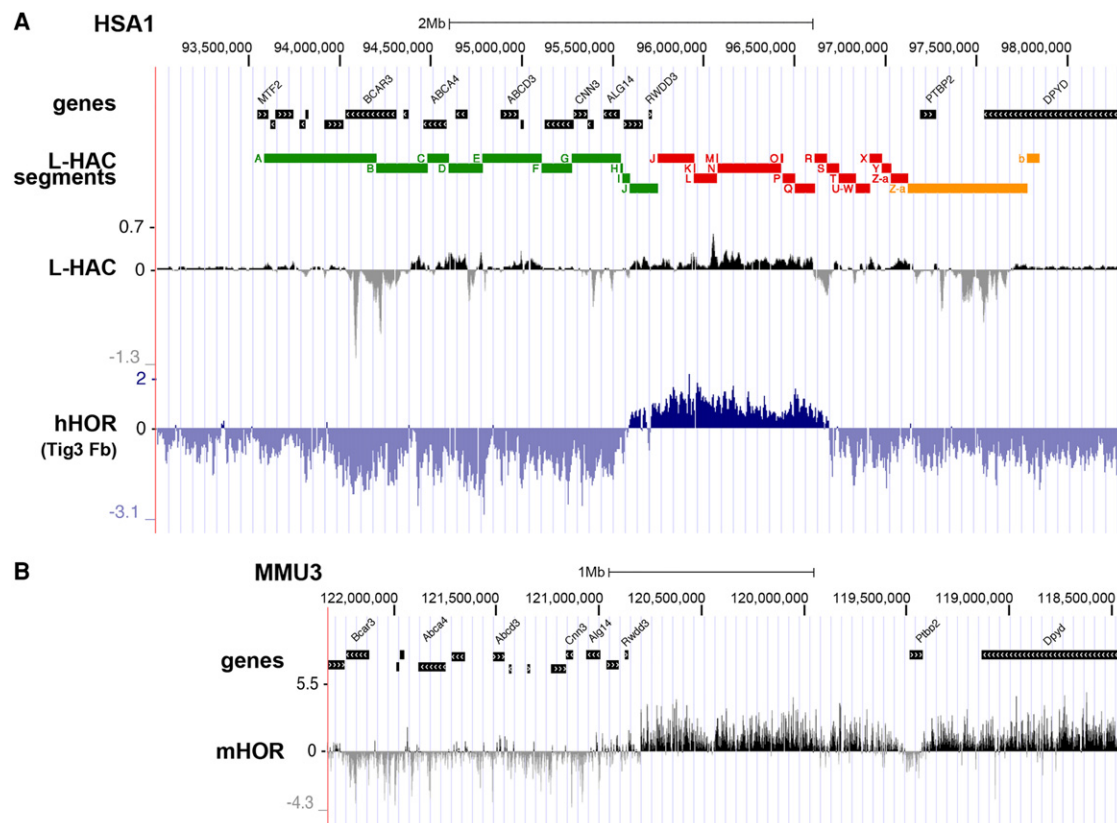

Figure 6. Nuclear lamina interactions of L-HAC in comparison to hHOR and mHOR. (*A*) DamID profile of nuclear lamina (NL) interactions (Dam-Lamin B1/Dam log$_2$ ratio) along the L-HAC in mouse A9 cells and along the corresponding region of human Chromosome 1 (HSA1) in human lung fibroblasts (Tig3 Fb; data from Guelen et al. 2008). (*B*) DamID profile along the mHOR on mouse Chromosome 3 (MMU3) in A9 cells carrying L-HAC. Annotation tracks are as in Figures 1 and 2.

on the native human and mouse chromosomes do not. In some instances, this could be due to sequence rearrangements in the HAC that may join these smaller regions to the gene-poor LAD and thus force them to be close to the NL (Supplemental Fig. S10). It is also possible that structural constraints in a native chromosomal context normally prevent these regions from contacting the NL. Interestingly, the gene-poor segment in the HAC shows no association with the NL, which is similar to the same region in human fibroblasts. However, the homologous mouse region (in the cells carrying the HAC) shows strong NL interactions (Fig. 6B). This suggests that this region in the human genome lacks a NL-targeting signal that is present in the corresponding mouse region.

From these results, we conclude that a large gene-poor LAD inside the linear HAC retains part of its ability to interact with the NL, despite disruption of the linear sequence by several structural rearrangements, its transfer to a different species, and its removal from a normal chromosomal context.

Discussion

In this work, we show that small, at least down to 0.6 Mb, chromosome segments spatially aggregate with similarly typed chromatin regions in a manner that is independent of their chromosomal context and the host genome that co-occupies the same nucleus. Therefore, segregation of the main chromatin classes in the nucleus appears to be an autonomous process.

By FISH and 4C-seq analyses, we confirmed previously established knowledge about spatial segregation of active and inactive chromatin domains (Simonis et al. 2006). First, gene-rich and gene-poor segments of both HACs segregate from each other in *cis*. We observed their distinct nuclear locations microscopically and further verified this finding by 4C-seq. Even small blocks of reshuffled EC and HC segments within HACs segregate from each other, suggesting that boundaries between active and inactive chromatin are stably maintained within HACs, probably by the same mechanisms responsible for boundary formation in endogenous chromosomes (Wang et al. 2014).

Second, we showed that distinct structural domains of HACs are differentially positioned in *trans* and associate with the nuclear compartments occupied by corresponding chromatin classes of the host genome. In the mouse host nuclei, gene-poor LINE/LTR-rich HAC segments spatially interact with LINE/LTR-rich HC but tend to avoid the EC environment. In contrast, gene-rich SINE-rich HAC segments tend to avoid contacts with HC but spatially interact with SINE-rich EC. Moreover, in both conventional and inverted nuclei, the human genes on the artificial chromosome are expressed according to their normal expression program, which is in agreement with previous data on proper transcriptional regulation of human genes within the mouse genome environment (Wilson et al. 2008). Finally, we show that human and mouse centromeres coalesce, forming common clusters of cHC, despite the fact that human α-satellite sequences and mouse minor satellite repeats are lacking homology and have only the 17-bp CENP-B box sequences in common (Choo 1997). In other words, HAC segments do not search for certain genomic sequences but rather for their preferred genomic environments built by certain chromatin classes.

Whereas the different chromatin classes properly segregate in the nucleus even if they are not contained within native chromosomes, nuclear lamina association is less autonomous and depends more on the chromosomal context. The major LAD within the HAC demonstrates weaker association with the NL compared to its chromosomal counterpart in human cells (Guelen et al. 2008)

or to the orthologous LAD in the mouse chromosome (Peric-Hupkes et al. 2010). Two reasons can account for this phenomenon. First, since the human centromere clusters with mouse centromeres on the chromocenter surface, the entire HAC is frequently pulled away from the NL where its gene desert can be labeled by DamID. Secondly, analysis of single-cell contact maps points to a cooperative LAD–NL interaction mechanism within each chromosome (Kind et al. 2015). Since there is only one major LAD in the HAC, such cooperativity is not expected.

Association of xenospecific chromatin segments found in this work points toward a mutual recognition of the same chromatin classes. It is well established that satellite sequences of centromeric and subcentromeric regions adhere to each other and cluster into chromocenters (Solovei et al. 2004a,b). Attraction between other repeats such as the SINEs and LINEs/LTRs that reside in EC and HC, respectively, could account for the segregation of other chromosomal regions (Krijger and de Laat 2013; Bouwman and de Laat 2015; Solovei et al. 2016). The orthologous segments of human and mouse chromosomes in this study are similar in their repeat repertoire. In both hHOR and mHOR, EC segments are enriched in SINEs, and HC segments are enriched in LINEs and LTRs (Supplemental Table S1). It seems that the SINEs/LINEs+LTRs ratio of a given chromosome segment is a parameter that can predict where in the interphase nucleus this segment will be situated (Supplemental Table S1; see also Meuleman et al. 2013). In agreement with our data, a correlation between 3D chromosome folding and enrichment of chromosome regions in similar interspersed repeats, SINEs in particular, was recently shown for human, mouse, and *Drosophila* genomes (Cournac et al. 2016).

The question of what mediates the mutual recognition of repeats remains open. The notion that colocalization of certain chromatin regions is due to sequestration by nuclear proteins, specifically recognizing DNA sequences, has been discussed, and in particular, it was suggested as a mechanism for ordered chromosome arrangement in elongated sperm nuclei (Joffe et al. 1998; Solovei et al. 1998). Clustering of loci of similar chromatin classes via protein binding is now increasingly suggested as the probable mechanism for most of the static and dynamic features of the nuclear architecture (Tang et al. 2015; Wijchers et al. 2016). Our data shed more light on this issue. In particular, the strong trend for colocalization of human and mouse centromere sequences can be explained by mutual recognition of centromere proteins building or associated with kinetochores (Masumoto et al. 1989; Muro et al. 1992). Previously, a similar recognition role was ascribed to the UBTF factor, driving localization of xenospecific or synthetic sequences to the nucleoli (Sullivan et al. 2001; Grob et al. 2014).

In case of interspersed repeats, the players remain unknown, and we can only speculate that they might be represented by such epigenetic cues as noncoding RNAs or chromatin proteins recognizing methylated DNA and posttranslational histone modifications. In accordance with this, recent studies suggest that recognition sequences for nuclear proteins can cause differential association of active chromatin with transcription factors, promoters and enhancers, and splicing machinery, and thus facilitate their clustering (Mitchell and Fraser 2008; Schoenfelder et al. 2010; Fanucchi et al. 2013, 2014; Tang et al. 2015). Similarly, clustering of inactive chromatin is mediated by associating with proteins involved in heterochromatin formation and maintenance, such as histone deacetylases, histone methyltransferases, DNA methylases, Polycomb group proteins, HP1, and many others (Denholtz

et al. 2013; Smith and Meissner 2013; Ciabrelli and Cavalli 2015; Schubeler 2015; Vieux-Rochas et al. 2015; Soshnev et al. 2016).

The ability of cells to establish a proper functional chromatin arrangement during early embryonic development and differentiation proves that the spatial organization of the nucleus is sufficiently encoded in the primary genomic sequence. However, the mechanisms translating linear genomic sequence into spatially ordered chromatin within the nucleus remain largely unknown. While the role of the nuclear lamina and nucleolus in chromatin scaffolding is well documented (Guelen et al. 2008; Nemeth et al. 2010; Peric-Hupkes et al. 2010; Solovei et al. 2013; Gonzalez-Sandoval et al. 2015; Talamas and Capelson 2015), the role of repetitive sequences, marking the major chromatin classes, is still poorly studied. The present work demonstrates that chromosome segments marked by a specific repeat repertoire autonomously position themselves within chromatin of their own class and therefore further supports our hypothesis about mutual recognition of repeats. We speculate that recognition of homotypic repetitive elements might be a possible mechanism driving the segregation of EC and HC, which is essential for establishing a functional nuclear architecture.

Methods

Animals, tissues, and cell lines

Animal protocols used in this study were approved by the Animal Ethic Committee of KU Leuven. For retina preparation, mice were killed by cervical dislocation according to the standard protocol. Eyes were enucleated immediately after death; retinas were excised and fixed with 4% formaldehyde in PBS for 12–24 h. Infiltration with sucrose, embedding, and cryosectioning were performed as described previously (Solovei 2010; Eberhart et al. 2012). Primary fibroblasts and A9 immortalized fibroblasts established from mice carrying HACs were grown in DMEM/F12 medium supplemented with 10% or 20% fetal calf serum. Cells with linear and circular HACs were grown with antibiotics for selection, blasticidin (6 µg/mL) and G418 (800 µg/mL), respectively. Human lung fibroblasts Tig3 were grown in DMEM, supplemented with 10% fetal calf serum and penicillin/streptomycin.

HAC sequencing

DNA for sequencing was isolated using the DNeasy Blood & Tissue kit (Qiagen) and sheared with the Bioruptor Next Gen UCD-300 sonicator (Diagenode). TruSeq DNA sequencing libraries (Illumina) were generated according to the manufacturer's instructions. The reads were aligned to the GRCh37 human reference genome, obtained from the UCSC Genome Browser by use of Burrows-Wheeler Alignment Tool (Li and Durbin 2009). Picard software (http://broadinstitute.github.io/picard/) was used to remove PCR duplicates from the resulting BAM files and to calculate the genomic coverages of hHOR (Chr 1: 93297341–97956405) in C-HAC and L-HAC. The depth of coverage was 5.8 and 3.9 for C-HAC and L-HAC respectively.

To determine the DNA copy number landscape of hHOR in C-HAC and L-HAC, the region was first divided into non-overlapping bins of 10.000 uniquely mappable positions similar to Baslan et al. (2016) and Moller et al. (2013). Subsequently, hHOR bins having a %GC-content lower than 28% were discarded, and the amount of sequence reads of C-HAC and L-HAC uniquely mapping in the remaining bins was counted. The log$R$-values for these bins were computed as the log$_2$ base of the read-count of a given bin divided by the average read-count of the bins located on hHOR, were cor-

rected for %GC-bias using a Loess fit in R (R Core Team 2016), and were finally normalized to the median of the hHOR log$R$-values. These log$R$-values were then segmented using circular binary segmentation (Olshen et al. 2004), and DNA-copy number was estimated as $2^{\log R} \times \Psi$, where Ψ denotes ploidy (Voet et al. 2013). Ψ was set to 1. To detect structural variants, we performed split-read and discordant read-pair analysis. In brief, paired-end sequences were mapped with BWA-MEM (http://arxiv.org/abs/1303.3997), enabling the detection of split reads. Since C-HAC segregates in a hamster cell line, the reads were mapped to a synthetic reference genome that consists of the human reference genome (GRCh37; hg19), the hamster reference genome (criGri1), and in-house nucleotide sequences of alphaSAT20 and of the constructs Neo-loxP-3′HPRT1, TEL08-cBAp-BLAS-loxP, and TEL08-SV40-DNA-5′HPRT-loxP. As the L-HAC passaged through both hamster and mouse genetic backgrounds, its reads were mapped to the same synthetic reference genome expanded with the mouse reference genome (mm10). Following mapping, PCR duplicates were removed using Picard, and split reads were identified. Additionally, discordantly mapped read-pairs, of which at least one read mapped to the hHOR, were selected and clustered. C-HAC and L-HAC discordant read-pairs were clustered if the reads in each pair mapped to the same chromosomes, the same strands, and within a distance of 500 bp. Clusters having at least two discordant read-pairs were considered potential structural variants. In addition, single discordant read-pairs corroborated by DNA copy number breakpoints in Integrative Genome Browser (IGV) were considered genuine structural variants ($n = 1$). Read-pairs mapping to the same class of repeat were discarded ($n = 16$). One discordant read-pair group was discarded since the reads did not map uniquely. All clusters of discordant read-pairs indicating a translocation between human and mouse, or between human and hamster sequences, could be excluded due to homology of mouse/hamster sequences to human sequence near the mate. The log$R$ values, segments, and chromosome rearrangement signatures were plotted by use of the Circos plot tool (Krzywinski et al. 2009).

Fluorescence in situ hybridization and confocal microscopy

BACs for human and mouse genomic loci used in the study are listed in the Supplemental Table S1 and were purchased from the BACPAC Resources Center (Children's Hospital Oakland Research Institute; http://bacpac.chori.org/). BAC DNA was amplified using the GenomiPhi kit (GE Healthcare) and labeled with dUTPs conjugated to various fluorochromes (FITC, Cy3, TexasRed, Cy5) by nick-translation (Cremer et al. 2007; Solovei 2010). Before hybridizing probes to cryosections, all purchased BAC clones were verified by cohybridizing with corresponding chromosome paints to metaphase chromosomes. FISH on cryosections was performed according to a protocol described elsewhere (Solovei 2010). For combined immuno-FISH, histone modifications were detected first, using primary and secondary antibodies; subsequently, sections were post-fixed with 2% formaldehyde and hybridized with DNA probes (Solovei 2010; Solovei and Cremer 2010). Primary antibodies were anti-H3K4me3 (Abcam, ab8580) and anti-H3K20me3 (kindly provided by A.H.F.M. Peters, Friedrich Meischer Institute for Biomedical Research, Switzerland). Stacks of optical sections through retina or cultured cells were collected using a Leica TCS SP5 confocal microscope equipped with Plan Apo 63×/1.4 NA oil immersion objective and lasers with excitation lines 405, 488, 561, 594, and 633 nm. Dedicated ImageJ plugins were used to compensate for axial chromatic shift between channels in confocal stacks and to create RGB stacks (Walter et al. 2006; Ronneberger et al. 2008).

Gene expression analysis

Tissues samples of mice carrying either C- or L-HACs and A9 cells carrying L-HAC were collected directly into lysis buffer, immediately frozen, and stored at −20°C. Isolation of RNA and reverse transcription were carried out as described previously (Szwagierczak et al. 2010). Five nanograms RNA were used for cDNA synthesis. Real-time PCR analysis was performed on the LightCycler 480 Instrument II (Roche) at standard reaction conditions using TaqMan Gene Expression Master Mix (Applied Biosystems). TaqMan assay IDs for the probes for human and mouse genes, respectively, were as following:

BCAR3/Bcar3 - Hs00981957_m1/Mm00600213_m1;
ABCA4/Abca4 - Hs00979594_m1/Hm00492035_m1;
ARHGAP29/Arhgap29 - Hs00191351_m1/Mm00805036_m1;
F3/F3 - Hs01076029_m1/Mm00438853_m1;
SLC44A3/Slc44a3 - Hs00537043_m1/Mm00520420_m1;
CNN3/Cnn3 - Hs01052674_m1/Mm00783337_s1;
ALG14/Alg14 - Hs00293655_m1/Mm00783337_s1;
TMEM56/Tmem56 - Hs00996279_m1/Mm00619261_m1;
PTBP2/Ptbp2 - HS00221842_m1/Mm00497922_m1;
GAPDH/Gapdh - Mm99999915_g1.

Gene expression levels were normalized to *GAPDH*/*Gapdh* and calculated using the comparative C_T method ($\Delta\Delta C_T$ method). Real-time analysis of A9 cells was performed in three biological replicates. Tissue-specific expression of *ABCA4*/*Abca4* and *CNN3*/*Cnn3* genes was analyzed by reverse transcription PCR using cDNA from pooled retina and liver samples from 10 mice of each mouse line carrying either L-HAC or C-HAC. The primers were designed across exon-exon boundaries and are listed in Supplemental Table S3.

4C-seq and analysis

Both the human (Tig3) and the mouse fibroblasts were grown in DMEM/10% FCS. Cells were cross-linked using 2% formaldehyde for 10 min at room temperature in 10% FCS/PBS. 4C templates were prepared as described previously (Splinter et al. 2011; van de Werken et al. 2012a). Chromatin was cross-linked with 2% formaldehyde in 10 mL PBS with 10% FCS for 10 min at room temperature, nuclei were isolated in 10 mL cold lysis buffer for 10 min, and cross-linked DNA was digested with HindIII. Digestion was followed by proximity ligation, removal of cross-links, a secondary restriction digestion with DpnII, and a second proximity ligation. Two hundred nanograms of the resulting 4C template were used for the subsequent PCR reaction using Expand Long Template Polymerase (Roche). The PCR products were purified using the High Pure PCR Product Purification kit (Roche) and the QIAquick PCR purification kit (Qiagen).

The primer sequences of the six viewpoints on the human chromosomal regions and the six viewpoints on the homologous chromosomal regions in mouse are shown in Supplemental Table S4. 4C-seq data analysis was performed as previously described (van de Werken et al. 2012a). In brief, the 4C-seq reading primer sequences were used to de-multiplex the 4C-seq samples, and the selected single-end reads were trimmed from the 5′ end to the 5′ end of the HindIII recognition site (AAGCTT). The trimmed reads were mapped, allowing no single mismatch, to a database of digested genome fragment-ends using the human reference genome build hg19 or the mouse reference genome build mm9 including the human HAC orthology region from build hg19. All the 4C-seq samples passed the common quality control threshold values (van de Werken et al. 2012a,b) except the gene-desert viewpoint from C-HAC, which was discarded (Supplemental Fig. S11). After mapping the 4C-seq reads, the 4C-seq data in *cis* was

smoothed using a running trimmed (10%) mean approach of 11 fragment-ends in a single window. 4C-seq *cis* profiles were visualized with the UCSC Genome Browser (Kent et al. 2002) and adapted manually. The 4C-seq *trans* data were first binarized and, in a running window of 500 fragment-ends, compared to the whole *trans* chromosome. *Trans*-interacting domains were called carrying out a binomial test and a multiple hypothesis testing correction using a false discovery rate (FDR) value <0.01 with permuted data (van de Werken et al. 2012a). These *trans*-interacting domains were plotted with the Circos plot tool (Krzywinski et al. 2009), and their GC, SINE, and LINE/LTR characteristics were calculated using GRanges (Lawrence et al. 2013) and RepBase (Bao et al. 2015). 4C-seq *cis* and *trans* profile comparisons were generated through a running mean approach with windows of 50 and 500 fragment-ends, respectively. Pairwise Spearman's rank correlation coefficients were calculated, discarding 200 kb around each viewpoints. The R statistical package was used for the statistical calculations and for generating the 4C-seq data plots (R Core Team 2016).

DamID analysis

DamID was performed as described (Guelen et al. 2008). Two independent replicate experiments were combined by averaging, after which a running median filter (window size 99 probes) was applied to reduce noise.

Data access

The 4C-seq data and DamID data from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) (Edgar et al. 2002) under accession numbers GSE84685 and GSE84952, respectively. The HAC DNA sequencing data have been submitted to the Sequence Read Archive (SRA; www.ncbi.nlm.nih.gov/sra) (Leinonen et al. 2011) under accession number SRP099474.

References

Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6:** 11.

Baslan T, Kendall J, Rodgers L, Cox H, Riggs M, Stepansky A, Troge J, Ravi K, Esposito D, Lakshmi B, et al. 2016. Corrigendum: genome-wide copy number analysis of single cells. *Nat Protoc* **11:** 616.

Bickmore WA, van Steensel B. 2013. Genome architecture: domain organization of interphase chromosomes. *Cell* **152:** 1270–1284.

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Muller S, Eils R, Cremer C, Speicher MR, et al. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* **3:** e157.

Bouwman BA, de Laat W. 2015. Getting the genome in shape: the formation of loops, domains and compartments. *Genome Biol* **16:** 154.

Choo AKH. 1997. *The centromere*. Oxford University Press, New York.

Ciabrelli F, Cavalli G. 2015. Chromatin-driven behavior of topologically associating domains. *J Mol Biol* **427:** 608–625.

Cournac A, Koszul R, Mozziconacci J. 2016. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res* **44:** 245–255.

Cremer T, Lichter P, Borden J, Ward DC, Manuelidis L. 1988. Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probes. *Hum Genet* **80:** 235–246.

Cremer M, Muller S, Kohler D, Brero A, Solovei I. 2007. Cell preparation and multicolor FISH in 3D preserved cultured mammalian cells. *CSH Protoc* **2007:** pdb prot4723.

Cremer T, Cremer C, Lichter P. 2014. Recollections of a scientific journey published in human genetics: from chromosome territories to interphase cytogenetics and comparative genome hybridization. *Hum Genet* **133:** 403–416.

Cremer T, Cremer M, Hubner B, Strickfaden H, Smeets D, Popken J, Sterr M, Markaki Y, Rippe K, Cremer C. 2015. The 4D nucleome: evidence for a dynamic nuclear landscape based on co-aligned active and inactive nuclear compartments. *FEBS Lett* **589:** 2931–2943.

Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* **145:** 1119–1131.

Denholtz M, Bonora G, Chronis C, Splinter E, de Laat W, Ernst J, Pellegrini M, Plath K. 2013. Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization. *Cell Stem Cell* **13:** 602–616.

Dixon JR, Gorkin DU, Ren B. 2016. Chromatin domains: the unit of chromosome organization. *Mol Cell* **62:** 668–680.

Eberhart A, Kimura H, Leonhardt H, Joffe B, Solovei I. 2012. Reliable detection of epigenetic histone marks and nuclear proteins in tissue cryosections. *Chromosome Res* **20:** 849–858.

Eberhart A, Feodorova Y, Song C, Wanner G, Kiseleva E, Furukawa T, Kimura H, Schotta G, Leonhardt H, Joffe B, et al. 2013. Epigenetics of eu- and heterochromatin in inverted and conventional nuclei from mouse retina. *Chromosome Res* **21:** 535–554.

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30:** 207–210.

Fanucchi S, Shibayama Y, Burd S, Weinberg MS, Mhlanga MM. 2013. Chromosomal contact permits transcription between coregulated genes. *Cell* **155:** 606–620.

Fanucchi S, Shibayama Y, Mhlanga MM. 2014. Are genes switched on when they kiss? *Nucleus* **5:** 103–112.

Goetze S, Mateos-Langerak J, Gierman HJ, de Leeuw W, Giromus O, Indemans MH, Koster J, Ondrej V, Versteeg R, van Driel R. 2007. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol* **27:** 4475–4487.

Gonzalez-Sandoval A, Towbin BD, Kalck V, Cabianca DS, Gaidatzis D, Hauer MH, Geng L, Wang L, Yang T, Wang X, et al. 2015. Perinuclear anchoring of H3K9-methylated chromatin stabilizes induced cell fate in *C. elegans* embryos. *Cell* **27:** 4475–4487.

Grob A, McStay B. 2014. Construction of synthetic nucleoli and what it tells us about propagation of sub-nuclear domains through cell division. *Cell Cycle* **13:** 2501–2508.

Grob A, Colleran C, McStay B. 2014. Construction of synthetic nucleoli in human cells reveals how a major functional nuclear domain is formed and propagated through cell division. *Genes Dev* **28:** 220–230.

Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al. 2008. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453:** 948–951.

Harr JC, Gonzalez-Sandoval A, Gasser SM. 2016. Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. *EMBO Rep* **17:** 139–155.

Jachowicz JW, Torres-Padilla ME. 2016. LINEs in mice: features, families, and potential roles in early development. *Chromosoma* **125:** 29–39.

Joffe BI, Solovei IV, Macgregor HC. 1998. Ordered arrangement and rearrangement of chromosomes during spermatogenesis in two species of planarians (*Plathelminthes*). *Chromosoma* **107:** 173–183.

Joffe B, Leonhardt H, Solovei I. 2010. Differentiation and large scale spatial organization of the genome. *Curr Opin Genet Dev* **20:** 562–569.

Kakeda M, Nagata K, Osawa K, Matsuno H, Hiratsuka M, Sano A, Okazaki A, Shitara S, Nishikawa S, Masuya A, et al. 2011. A new chromosome 14-based human artificial chromosome (HAC) vector system for efficient transgene expression in human primary cells. *Biochem Biophys Res Commun* **415:** 439–444.

Kazuki Y, Kobayashi K, Aueviriyavit S, Oshima T, Kuroiwa Y, Tsukazaki Y, Senda N, Kawakami H, Ohtsuki S, Abe S, et al. 2013. Trans-chromosomic

mice containing a human CYP3A cluster for prediction of xenobiotic metabolism in humans. *Hum Mol Genet* **22:** 578–592.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12:** 996–1006.

Kind J, Pagie L, de Vries SS, Nahidiazar L, Dey SS, Bienko M, Zhan Y, Lajoie B, de Graaf CA, Amendola M, et al. 2015. Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163:** 134–147.

Krijger PH, de Laat W. 2013. Identical cells with different 3D genomes; cause and consequences? *Curr Opin Genet Dev* **23:** 191–196.

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. *Genome Res* **19:** 1639–1645.

Kupper K, Kolbl A, Biener D, Dittrich S, von Hase J, Thormeyer T, Fiegler H, Carter NP, Speicher MR, Cremer T, et al. 2007. Radial chromatin positioning is shaped by local gene density, not by gene expression. *Chromosoma* **116:** 285–306.

Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ. 2013. Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9:** e1003118.

Leinonen R, Sugawara H, Shumway M, International Nucleotide Sequence Database Consortium. 2011. The sequence read archive. *Nucleic Acids Res* **39:** D19–D21.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25:** 1754–1760.

Lichter P, Cremer T, Borden J, Manuelidis L, Ward DC. 1988. Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant DNA libraries. *Hum Genet* **80:** 224–234.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326:** 289–293.

Markaki Y, Smeets D, Cremer M, Schermelleh L. 2013. Fluorescence in situ hybridization applications for super-resolution 3D structured illumination microscopy. *Methods Mol Biol* **950:** 43–64.

Masumoto H, Masukata H, Muro Y, Nozaki N, Okazaki T. 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol* **109:** 1963–1973.

Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, Reinders M, Wessels L, van Steensel B. 2013. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23:** 270–280.

Mitchell JA, Fraser P. 2008. Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev* **22:** 20–25.

Moller EK, Kumar P, Voet T, Peterson A, Van Loo P, Mathiesen RR, Fjelldal R, Grundstad J, Borgen E, Baumbusch LO, et al. 2013. Next-generation sequencing of disseminated tumor cells. *Front Oncol* **3:** 320.

Muro Y, Masumoto H, Yoda K, Nozaki N, Ohashi M, Okazaki T. 1992. Centromere protein B assembles human centromeric α-satellite DNA at the 17-bp sequence, CENP-B box. *J Cell Biol* **116:** 585–596.

Naumova N, Dekker J. 2010. Integrating one-dimensional and three-dimensional maps of genomes. *J Cell Sci* **123:** 1979–1988.

Nemeth A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, Solovei I, Cremer T, Dopazo J, Langst G. 2010. Initial genomics of the human nucleolus. *PLoS Genet* **6:** e1000889.

O'Doherty A, Ruf S, Mulligan C, Hildreth V, Errington ML, Cooke S, Sesay A, Modino S, Vanes L, Hernandez D, et al. 2005. An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science* **309:** 2033–2037.

Olshen AB, Venkatraman ES, Lucito R, Wigler M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5:** 557–572.

Oshimura M, Uno N, Kazuki Y, Katoh M, Inoue T. 2015. A pathway from chromosome transfer to engineering resulting in human and mouse artificial chromosomes for a variety of applications to bio-medical challenges. *Chromosome Res* **23:** 111–133.

Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Graf S, Flicek P, Kerkhoven RM, van Lohuizen M, et al. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38:** 603–613.

Pueschel R, Coraggio F, Meister P. 2016. From single genes to entire genomes: the search for a function of nuclear organization. *Development* **143:** 910–923.

R Core Team. 2016. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the

human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159:** 1665–1680.

Ronneberger O, Baddeley D, Scheipl F, Verveer PJ, Burkhardt H, Cremer C, Fahrmeir L, Cremer T, Joffe B. 2008. Spatial quantitative analysis of fluorescently labeled nuclear structures: problems, methods, pitfalls. *Chromosome Res* **16:** 523–562.

Schermelleh L, Heintzmann R, Leonhardt H. 2010. A guide to super-resolution fluorescence microscopy. *J Cell Biol* **190:** 165–175.

Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS, et al. 2010. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42:** 53–61.

Schubeler D. 2015. Function and information content of DNA methylation. *Nature* **517:** 321–326.

Sexton T, Cavalli G. 2015. The role of chromosome domains in shaping the functional genome. *Cell* **160:** 1049–1059.

Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38:** 1348–1354.

Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14:** 204–220.

Solovei I. 2010. Fluorescence in situ hybridization (FISH) on tissue cryosections. *Methods Mol Biol* **659:** 71–82.

Solovei I, Cremer M. 2010. 3D-FISH on cultured cells combined with immunostaining. *Methods Mol Biol* **659:** 117–126.

Solovei IV, Joffe BI, Hori T, Thomson P, Mizuno S, Macgregor HC. 1998. Unordered arrangement of chromosomes in the nuclei of chicken spermatozoa. *Chromosoma* **107:** 184–188.

Solovei I, Grandi N, Knoth R, Volk B, Cremer T. 2004a. Positional changes of pericentromeric heterochromatin and nucleoli in postmitotic Purkinje cells during murine cerebellum development. *Cytogenet Genome Res* **105:** 302–310.

Solovei I, Schermelleh L, During K, Engelhardt A, Stein S, Cremer C, Cremer T. 2004b. Differences in centromere positioning of cycling and postmitotic human cell types. *Chromosoma* **112:** 410–423.

Solovei I, Kreysing M, Lanctot C, Kosem S, Peichl L, Cremer T, Guck J, Joffe B. 2009. Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137:** 356–368.

Solovei I, Wang AS, Thanisch K, Schmidt CS, Krebs S, Zwerger M, Cohen TV, Devys D, Foisner R, Peichl L, et al. 2013. LBR and lamin A/C sequentially tether peripheral heterochromatin and inversely regulate differentiation. *Cell* **152:** 584–598.

Solovei I, Thanisch K, Feodorova Y. 2016. How to rule the nucleus: divide et impera. *Curr Opin Cell Biol* **40:** 47–59.

Soshnev AA, Josefowicz SZ, Allis CD. 2016. Greater than the sum of parts: complexity of the dynamic epigenome. *Mol Cell* **62:** 681–694.

Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, Zhu Y, Kaaij LJ, van Ijcken W, Gribnau J, Heard E, et al. 2011. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev* **25:** 1371–1383.

Stults DM, Killen MW, Pierce HH, Pierce AJ. 2008. Genomic architecture and inheritance of human ribosomal RNA gene clusters. *Genome Res* **18:** 13–18.

Sullivan GJ, Bridger JM, Cuthbert AP, Newbold RF, Bickmore WA, McStay B. 2001. Human acrocentric chromosomes with transcriptionally silent nucleolar organizer regions associate with nucleoli. *EMBO J* **20:** 2867–2874.

Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* **38:** e181.

Talamas JA, Capelson M. 2015. Nuclear envelope and genome interactions in cell fate. *Front Genet* **6:** 95.

Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J, Ruszczycki B, et al. 2015. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163:** 1611–1627.

Tomizuka K, Yoshida H, Uejima H, Kugoh H, Sato K, Ohguma A, Hayasaka M, Hanaoka K, Oshimura M, Ishida I. 1997. Functional expression and germline transmission of a human chromosome fragment in chimaeric mice. *Nat Genet* **16:** 133–143.

Umbarger MA, Toro E, Wright MA, Porreca GJ, Bau D, Hong SH, Fero MJ, Zhu LJ, Marti-Renom MA, McAdams HH, et al. 2011. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol Cell* **44:** 252–264.

van de Werken HJ, de Vree PJ, Splinter E, Holwerda SJ, Klous P, de Wit E, de Laat W. 2012a. 4C technology: protocols and data analysis. *Methods Enzymol* **513:** 89–112.

van de Werken HJ, Landan G, Holwerda SJ, Hoichman M, Klous P, Chachik R, Splinter E, Valdes-Quezada C, Oz Y, Bouwman BA, et al. 2012b. Robust 4C-seq data analysis to screen for regulatory DNA interactions. *Nat Methods* **9:** 969–972.

van Steensel B, Delrow J, Henikoff S. 2001. Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* **27:** 304–308.

Vieux-Rochas M, Fabre PJ, Leleu M, Duboule D, Noordermeer D. 2015. Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc Natl Acad Sci* **112:** 4672–4677.

Voet T, Vermeesch J, Carens A, Durr J, Labaere C, Duhamel H, David G, Marynen P. 2001. Efficient male and female germline transmission of a human chromosomal vector in mice. *Genome Res* **11:** 124–136.

Voet T, Schoenmakers E, Carpentier S, Labaere C, Marynen P. 2003. Controlled transgene dosage and PAC-mediated transgenesis in mice using a chromosomal vector. *Genomics* **82:** 596–605.

Voet T, Kumar P, Van Loo P, Cooke SL, Marshall J, Lin ML, Zamani Esteki M, Van der Aa N, Mateiu L, McBride DJ, et al. 2013. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. *Nucleic Acids Res* **41:** 6119–6138.

Walter J, Joffe B, Bolzer A, Albiez H, Benedetti PA, Muller S, Speicher MR, Cremer T, Cremer M, Solovei I. 2006. Towards many colors in FISH on 3D-preserved interphase nuclei. *Cytogenet Genome Res* **114:** 367–378.

Wang J, Lawry ST, Cohen AL, Jia S. 2014. Chromosome boundary elements and regulation of heterochromatin spreading. *Cell Mol Life Sci* **71:** 4841–4852.

Weierich C, Brero A, Stein S, von Hase J, Cremer C, Cremer T, Solovei I. 2003. Three-dimensional arrangements of centromeres and telomeres in nuclei of human and murine lymphocytes. *Chromosome Res* **11:** 485–502.

Weuts A, Voet T, Verbeeck J, Lambrechts N, Wirix E, Schoonjans L, Danloy S, Marynen P, Froyen G. 2012. Telomere length homeostasis and telomere position effect on a linear human artificial chromosome are dictated by the genetic background. *Nucleic Acids Res* **40:** 11477–11489.

Wijchers PJ, Krijger PH, Geeven G, Zhu Y, Denker A, Verstegen MJ, Valdes-Quezada C, Vermeulen C, Janssen M, Teunissen H, et al. 2016. Cause and consequence of tethering a subTAD to different nuclear compartments. *Mol Cell* **61:** 461–473.

Wilson MD, Barbosa-Morais NL, Schmidt D, Conboy CM, Vanes L, Tybulewicz VL, Fisher EM, Tavare S, Odom DT. 2008. Species-specific transcription in mice carrying human chromosome 21. *Science* **322:** 434–438.

Small chromosomal regions position themselves autonomously according to their chromatin class

Harmen J.G. van de Werken, Josien C. Haan, Yana Feodorova, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2017/04/26/gr.213751.116.DC1
References	This article cites 90 articles, 21 of which can be accessed free at: http://genome.cshlp.org/content/27/6/922.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**