Factors shaping the 3D genome

Yun Zhu

ISBN-978-90-393-6366-9

Layout and Printing: Off Page

Cover: Picture was provided by Dreamstime

Factors shaping the 3D genome

Factoren die de ruimtelijke vouwing van het genoom bepalen

(met een samenvatting in het Nederlands)

Proefschrift

Ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. G.J. van der Zwaan, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 7 Juli 2015 des middags te 12.45 uur

door

Yun Zhu

geboren op 4 november 1983 te Nanyang, China

Promotor: **Prof. dr. W.L. de laat**

The research described in this thesis was performed at the Hubrecht Institute Of the Royal Dutch Academy of Arts and Science (KNAW).
This thesis was (partly) accomplished with financial support from a European Research Council Starting Grant (209700, '4C') and an NWO VICI grant 724.012.003.

TABLE OF CONTENTS

SCOPE OF THIS THESIS

The shape of the genome has emerged as an important regulator of transcription. With the development of chromosome conformation capture (3C) related techniques and with sequencing depth increasing we are getting an ever improving picture of the architectural fine-structure of the genome. There are many published examples of individual genes that alter their nuclear location in relation to changes in their expression status. However, the mechanisms behind spatial re-positioning, and whether there is any causal relationship between gene location and expression, are still unclear. The work described in this thesis aimed to decipher this relationship in more detail and sought to unravel the role of a number of key transcription factors and chromatin proteins in the 3D organization of the genome.

Chapter 1 gives a general introduction on the spatial organization of the genome, the factors shaping it and its functional significance. Tethering systems that enable manipulating chromosome folding, some of which were used in this work, are introduced.

Chapter 2 describes the application of LacO/LacR platform which we used to study the impact of tethering different types of repressive chromatin regulators to transcriptionally active genomic loci. We found that susceptibility to nuclear repositioning depends on the genomic context of a locus and that the degree and directionality of repositioning depends on the type of factor recruited. Chromosomal regions with similarly typed chromatin preferentially cluster together, which we speculate may contribute to maintenance of the respective chromatin states. Importantly, spatial repositioning is not causally related to gene expression changes. Thus, where local chromatin loops between enhancers and promoters clearly impact on gene expression, our data support accumulating evidence that higher-order levels of nuclear organization are not a major determinant of gene expression.

Chapter 3 demonstrates a number of unique features of the pluripotent 3D genome. We demonstrate that particularly inactive chromatin is unusually un-organized in embryonic stem cells and we show that pluripotency factors shape the active chromatin compartment in the pluripotent cell nucleus by bringing together the chromosomal regions with high density of binding sites for these factors. We speculate this organization may facilitate robust expression of key pluripotency genes and thereby contribute to maintenance of the pluripotent state.

Linker histone H1 is studied in **Chapter 4**. Here, we generated and combined various genome-wide data sets to investigate the role of histone H1 in chromatin folding and function. Our data shows that cells require normal histone H1 levels to expose their proper regulatory landscape, indicate that histone H1 is important for the topological segmentation of chromosomes and suggests that topological alterations are more related to epigenetic than to transcriptional changes.

In **chapter 5**, our findings are discussed and put in perspective of existing models for genome organization.

1

Introduction

For decades already, scientists are trying to find out the composition of our genome and how the epigenetic code influences its function. In the early 90s, the human genome project was initiated to identify and map all the genes, and it was realized that all the coding parts together only cover a few percentages of the whole genome [1]. In 2003, the Encyclopedia of DNA Elements (ENCODE) was launched to identify the regulatory parts of our genome in both the coding and non-coding regions (Fig.1)[2] . The non-coding genome, previously thought of as junk DNA, is now appreciated to be full of regulatory sequences, including promoters and enhancers, which critically control the expression of genes. They enable that many genes can be expressed in a tissue-and development-specific manner, a requirement for cells to adapt to environmental stress and for the evolutionary rise of multi-cellular organisms. Studying how those elements act to control the expression of nearby and distal genes is one of the current major challenges in the field of functional genomics.

CHROMATIN AND TRANSCRIPTION

Every tiny cell nucleus in our body measures approximately 10 micrometer across and contains, if unfolded, roughly 2 meters of DNA. DNA has to be compacted to fit in such a small cellular compartment. Therefore, DNA is packed into chromatin, a protein scaffold that causes DNA condensation and hampers access of other regulatory factors to the DNA template. The basic chromatin compaction unit is the nucleosome, which is composed of an octamer of four core histone proteins with ~147 basepair (bp) of DNA wrapped, in two turns, around the particle. In between the particles are small stretches of linker DNA, which collectively, under the electron microscope, looks like "beads on a string". Histone H1 can bind to linker DNA and thereby affect nucleosome position, which can promote and stabilize the folding of chromatin (further discussed in Chapter 4). In vivo, the "beads on a string" structure is further folded into higher-order structure of which the exact composition is still under debate and likely to differ along the chromatin fiber in a manner dependent on the locally associated proteins and RNA.

Chromatin folding and compaction is further dictated by the trans-acting factors that associate with DNA, with differential protein and RNA recruitment across the genome predominantly regulated at the level of the primary DNA sequence. DNase I hypersensitive sites (DHSs) are genomic sites where DNA is exposed and accessible to nuclease digestion by the enzyme DNaseI; these sites typically represent the regulatory DNA regions. They include promoter regions, where transcription is initiated, but most of DHSs are distal from transcription start sites (TSS) in non-coding genome, and function as enhancers and locus control regions [3, 4]. Histones also provide a rich platform for epigenetic marking. They can be acetylated, methylated, phosphorylated, ubiquitinated or otherwise chemically modified at various amino acid residues, with such often tissue- and site-specific modifications serving for the attraction or repulsion of specific trans-acting factors.

The same DNA sequence exists in very different cell types: how does a cell ensure that its specific repertoire of genes is being used? Enhancer and promoter interactions are essential for this tissue specific gene regulation in development. Enhancers are short DNA sequences

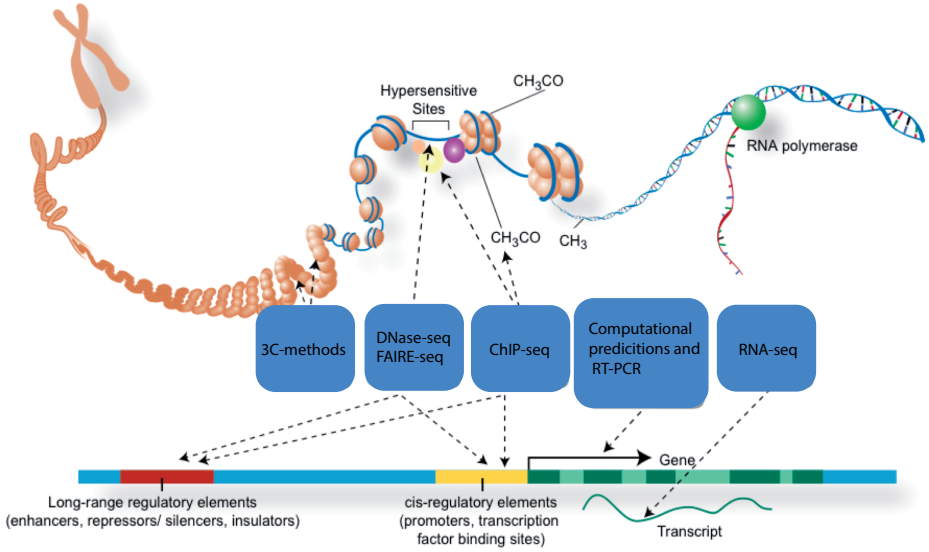


Figure 1. Functional genomic elements being identified by the ENCODE project. Adapted from [2].

embedded in the non-coding genome. They can act in both directions of genes, and affect genes over distance (Fig.2). With the development of high-throughput methods, it has been suggested that millions of putative enhancers may exist in our genome [5]. To understand the functional relationship between regulatory elements and genes, the three-dimensional (3D) structure of genome is an important aspect to study.

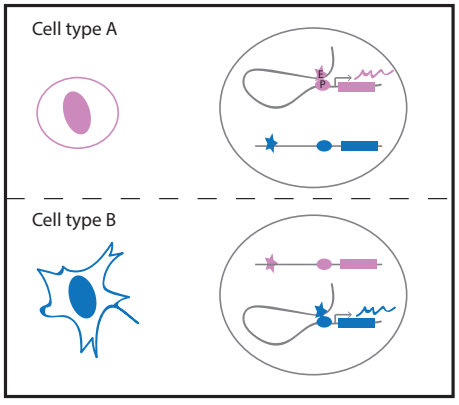


Figure 2. Enhancers regulate the expression of cell type specific genes. Different enhancers (E) interact with promoters (P) to activate different genes in cell type A and B.

SPATIAL ORGANIZATION OF THE GENOME

Heterochromatin and Euchromatin

Chromatin is not randomly distributed in the nucleus. One subdivision of compaction and folding is already appreciable when stained DNA is studied under the microscope. Two types of chromatin have been revealed by light microscope in the interphase nucleus, according to the differential stainings and compaction patterns. Heterochromatin marks darkly stained and highly condensed chromatin, while euchromatin describes the lightly stained and more open chromatin parts of the genome. Later studies suggested that heterochromatin provides a transcriptional inactive environment, with its extremely compact structure being less accessible for RNA polymerase and regulatory factors. In mammalian cells, heterochromatin is subdivided into two forms: constitutive and facultative heterochromatin. Constitutive heterochromatin is a relatively stable, tissue-invariant structure formed at centromeres, telomeres and repetitive sequences. These sequence elements play an important role in genome stability. Facultative heterochromatin is more flexible and has the ability to transmit to an open and active chromatin structure in a cell-type dependent manner [6, 7]. Euchromatin, as mentioned, refers to the more open chromatin parts of the genome. These are usually the more gene dense regions where active transcription takes place (Fig.3).

Chromosome Territory

Another major feature of chromatin structure is the existence of chromosome territories (CT). During mitosis, each chromosome becomes highly condensed. Decondensation takes place in the early G1 phase of the cell cycle but to an extent that chromosomes still occupy their own nuclear space during the rest of the cell cycle, known as CT. In situ hybridization studies discover that CTs are not randomly scattered in the nuclear space.

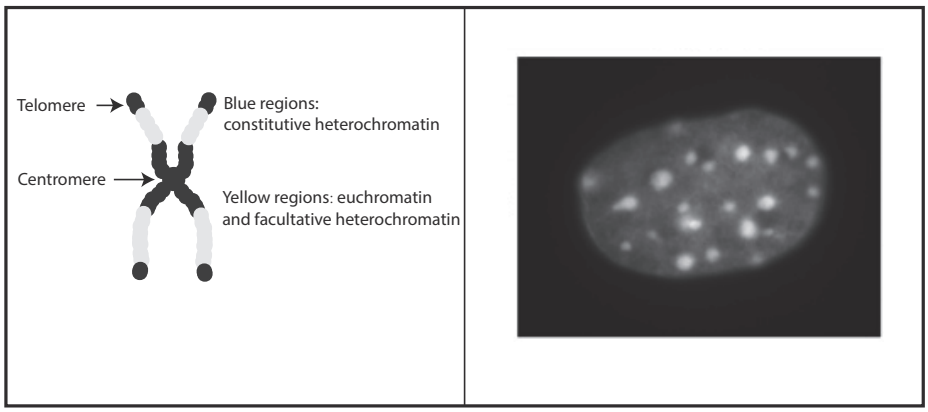


Figure 3. Heterochroamtin and Euchromatin. Left: In mitotic chromosomes, constitutive heterochromatin is seen around telomere, centromere and satellite repeats; Right: In interphase nuclei, heterochromatin can be visualized as DAPI-dense spots, mouse NIH 3T3 cell [100].

In human lymphocyte nuclei, chromosomes 18 and 19 have similar chromosome size, but their morphology and nuclear locations are quite different. Chromosome 19, containing gene dense regions and early replicating domains, predominantly shows an internal position. Chromosome 18, which has lots of gene deserts and mid-replicate domains, is often found at the nuclear periphery [8, 9]. That position variation related to gene density is further supported by analyzing radial positions of multiple gene clusters on various chromosomes in multiple cell types [10]. In addition to gene density, size matters. Larger chromosomes prefer to adopt peripheral positions while smaller chromosomes tend to be found in the interior of the nucleus [11]. Thus, CTs show a non-random distribution, but exact predictions about their location in a given cell are difficult to make, as positioning mostly follows probabilistic rules. CTs are also found to intermingle with each other by using fluorescence in situ hybridization (FISH) on ultrathin cryosections (cryoFISH) [12], which is a more sensitive and less destructive method than normal FISH. Those intermingled regions are involved in inter-chromosome interactions, and often show high potential to be translocation partners [13]. Most of these intermingled regions are associated with active transcripts. The inhibition of ongoing transcription only marginally affects CT intermingling, suggesting that inter-chromosomal contacts not necessarily rely on transcription [12].

Gene position in the nucleus

Radial distribution of individual gene in the nucleus appears also not randomly, but the exact principles of gene positioning are still under debate. The relative distance between genes to CT and other nuclear subcompartments is correlated with gene activity. Highly transcribed gene clusters prefer to locate outside their own CT and it has been suggested that such external position enables better assess of RNA polymerase and other regulatory factors in favor of efficient transcription. In contrast, the interior of CTs appears to retain inactive genes. Here, the compact chromatin may form a unique compartment separated from the rest, with a high concentration of silencing factors [14-16]. Our understanding of gene positioning relative to CTs comes from microscopy studies and although in these studies many genes seem to adhere to the rule that their activity determines their CT location. The idea of inactive genes being buried inside CTs is currently hard to reconcile with genomics approaches that show that inactive genes also preferentially line up at the peripheral nuclear lamina (discussed below). When genes change expression, they are sometimes seen to also move their positions in the nucleus. A few studies for example suggest that gene clusters can loop out of their CT upon transcription activation [17]. Similarly, it has been described that genes adopt different positions in the cell nucleus in different cell types, in relation with the changes in their transcriptional status [18-21]. For example, during embryonic cell (ES) differentiation, Mash1 gene is switched on and upregulated more than 100-fold. This strong activation is companied by this locus moving from the nuclear periphery to the interior of the nucleus [20]. A same cell-type and transcription-dependent switch in nuclear positioning was previously suggested for the B-cell specific immunoglobulin heavy chain (IgH) locus [22], but a careful allele-specific analysis of the genomic environment of IgH in various cell types could not provide support for this concept [23]. In cases when repositioning occurs, it is still unclear whether this is a cause or consequence of a transcriptional change.

In the case of Mash1 locus, the nuclear position of two neighboring genes is also dragged along from the nuclear periphery to the interior part. However, they remain silent during the process of reposition. Experimental inhibition of ongoing transcription has very little impact on the positioning of genes, indicating that maintenance of a nuclear location does not depend on transcription or engagement with RNA polymerase II (RNA polII). Local chromatin de-compaction is also observed during gene reposition. For example, during ES cell differentiation, the Hoxd cluster gets activated which is accompanied by chromatin unfolding and looping out of CT[24]. Interestingly, chromatin de-compaction happens before gene activation and repositioning, suggesting that 'openness' of chromatin, more than transcriptional activity, may determine where a gene or a gene locus ends up in the nuclear space [25].

3D structure and 3C technologies

Most of the findings discussed above are based on microscopy observations by applying FISH. FISH is one of the traditional technologies to investigate the 3D folding of the genome. However, FISH technology has its limitations. Only low numbers of loci can be visualized simultaneously in relatively low numbers of cells. Novel genomics methods based on proximity capture have developed quickly over the last ten years (Fig.4). Chromosome conformation capture (3C) was first developed for applications in yeast and then modified for mammalian system [26, 27]. 3C based methods often use formaldehyde to crosslink the chromatin and create a snapshot of the proximity of different parts of the genome in three dimensions within a population of cells. The cross-linked chromatin is then digested with restriction enzymes recognizing 4bp sequences (4bp-cutter) or 6pb sequences (6bp-cutter), and ligated in diluted conditions. Contact frequencies are then measured by determining the frequency of ligation events. In 3C technology, (semi-)quantitative PCR is used to evaluate the contact frequency of two known parts. The β-globin locus is one of the best studied loci by 3C. Only when active, the globin genes show specific contacts with a series of upstream enhancers that collectively form a so-called locus control region (LCR), as well as with other regulatory elements surrounding the locus [27, 28]. These regulatory long-range contacts depend on transcription factors binding to the corresponding enhancers and gene promoters [29-31]. Chromosome conformation capture-on-chip (4C) is designed as a "one versus all" strategy, as it can identify all the interacting partners of a chosen locus across the genome by using microarrays or next-generation sequencing (NGS) to measure the contact sequences [32, 33]. The first 4C study revealed that genes prefer to contact different regions across the remainder of their chromosomes and on other chromosomes in a manner that depends on their expression status. Thus, the active β-globin locus contacts many active regions not only on its own chromosomes, but also on other chromosomes. When the locus is inactive, it interacts with inactive regions, suggesting there is a spatial separation between active and inactive chromatin. 4C technology focuses on revealing the spatial contacts of individual genomic sites, and turns out to be a powerful assay to identify the structural role of transcription factors and regulatory elements. Many 4C studies suggest cell specific transcription factors are critical for setting up unique local or global chromatin structure, which will be discussed in detail later [33-36]. Enhancers and promoters are actively engaged

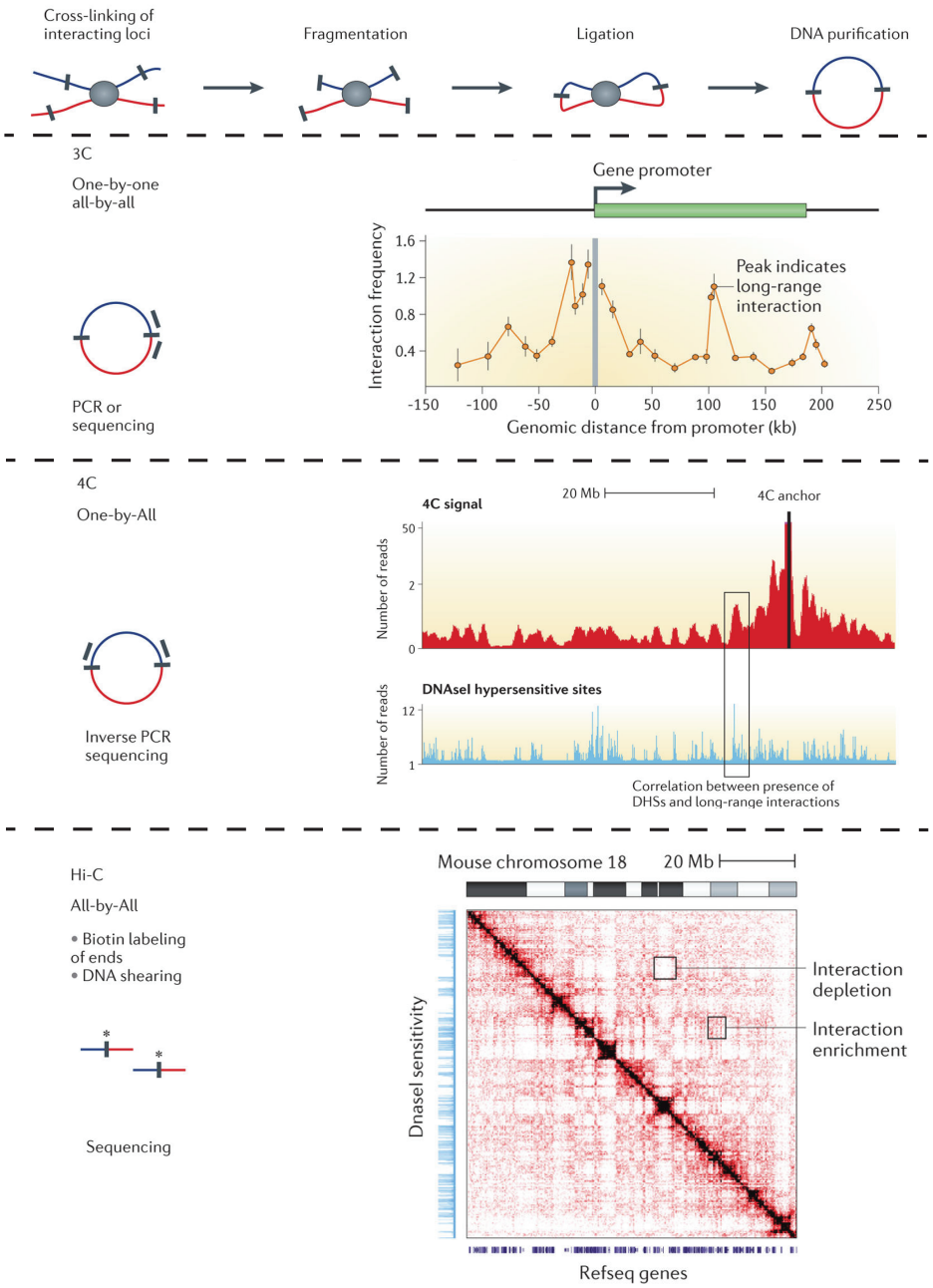


Figure 4. Overview of 3C related technologies and corresponding datasets. The upper panel shows the same steps in all the 3C related technologies. 3C data is for the CFTR gene in Caco2 cells [101]; 4C and Hi-C data are from [102]. Adapted from [103].

in extensive long-range interactions in a cell-type-specific manner [37], emerging as an important aspect of gene regulation. The high resolution of 4C also enables it as a robust method to identify enhancer and promoter interactions [38].

Yet another 3C-type method is Hi-C, which has been developed as an "all versus all" method that aims to map all the interactions in the genome. Initial Hi-C studies revealed that the 3D genome is separated into two compartments, which was also suggested by previous 4C studies. One spatial compartment possesses chromosomal regions with open chromatin, active genes and many DHSs, while the other compartment mainly includes gene deserts with inactive histone modifications, lacking DHSs [39]. As the sequencing depth increased, Hi-C can offer a more comprehensive and detailed map of genome structure. Dixon *et al* discovered that megabase-sized domain structures exist in each compartment, named Topologically Associating Domains, or "TADs" [40]. TADs are spatially separated domains, and serve as basic blocks of the genome structure. The boundaries between TADs are enriched for insulator binding protein CCCTC-binding factor (CTCF), housekeeping genes, short interspersed nuclear elements (SINE), and transfer RNAs (tRNAs), which may function as a barrier in setting up the genome structure. Surprisingly, TAD organization is stable across different cell types and highly conserved over evolution. The mechanism defining those TADs is still unclear. A recent study showed that deleting a boundary region between two TADs caused increased contacts between the domains and misregulation of their genes [41]. Overall, 3C based methods significantly enrich our knowledge of higher order chromatin structure. A caveat of the methods though is that they provide cell population-based averages of structures without distinguishing between cell-to-cell variations. Recently, Hi-C was applied in single cell. This study confirmed that domain structures exist in single nucleus, and also discovered that at higher-order levels of organization huge variation on chromosome topology exists between different cells [42]. Now, the urgent desire is to combine single cell structure analysis with single cell RNA sequencing, in order to better understand the functional role of chromatin structure in gene regulation.

Lamina-Associated Domains

Inside the nucleus, the inner nuclear membrane is coated with a network of filamentous lamin proteins, a structure that is known as the nuclear lamina. The peripheral nuclear space that borders the nuclear lamina mostly contains densely stained heterochromatin [43]. Nuclear lamina is not only important for the structure of the nuclear membrane, but also takes an active role in gene regulation. Mutations of any of the lamin proteins can impair the development of cells, and some of the mutations lead to gene position and expression changes [44-47]. Lately, the application of a high throughput method called DNA adenine methyltransferase identification (Dam-ID) used for the identification of genomic parts associated with nuclear lamina proteins led to the systemic uncovering of nuclear lamina associated domains, or "LADs" [48, 49]. In mammalian cells, LADs constitute 35%-40% of the genome, and are scattered across all the chromosomes with different densities (Fig.5) [50]. In general, LADs have lower gene density and possess distinct inactive signatures, with RNA polII and active histone modifications being depleted. During mouse ES cell differentiation towards neuronal lineages, LADs cumulatively change and show a cell type

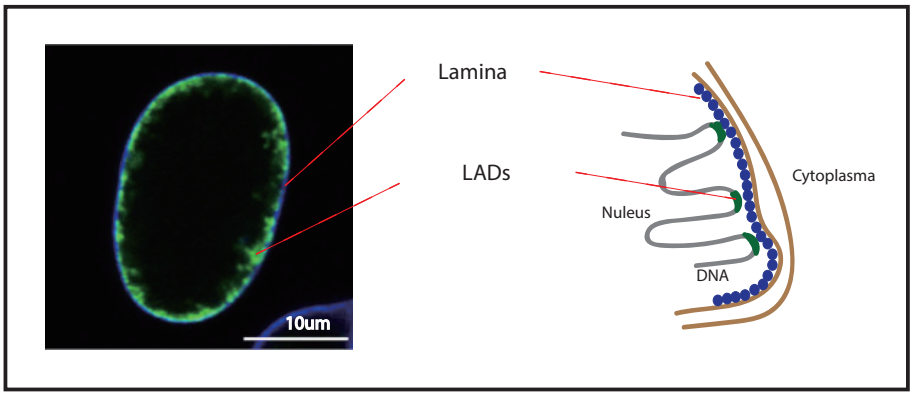


Figure 5. Lamina-Associated Domain. LADs (green) could be traced in real-time by ^{m6}A-tracer system. LaminB staining is shown in blue [53].

specific pattern. Some of the genes that get activated or poised for activation during lineage commitment lose their contacts with the nuclear lamina, while some of the genes that become silenced appear as new LADs in differentiated cells, suggesting that association with lamina is important for gene regulation in development [49]. How LADs are established is still unclear. Sequence context appears important as LADs have a distinctly increased A/T content [51, 52] yet this does not explain why during development LAD organization changes significantly. Recently, a single cell study revealed that only ~30% of LADs interact at any given time with the nuclear periphery and this close association is not stably inherited over cell cycles. After cell division, LADs stochastically relocate. H3K9 methyltransferase G9a has been proven to have the ability to promote nuclear periphery contacts, and the G9a induced histone modification H3K9me2 is also found to be enriched at the periphery and in LADs, indicating that this histone modification may play a role in determining the position of LADs [53]. However, the mechanism of LAD's dynamic association and dissociation with the nuclear periphery needs further investigation. A way to study the role of nuclear lamina in gene regulation involves the induced tethering of genomic loci to the nuclear periphery, which will be discussed in a later part of this introduction.

FACTORS SHAPING THE GENOME

Transcription factories

The concept of "transcription factory" was raised in early days by the visualization of nuclear sites of active transcription and of RNA polII molecules by immunofluorescence [54, 55]. Microscopy studies indicate that transcription is not uniformly spread over the nucleus and only a limited number of RNA polII foci appear to exist, known as "transcription factories" [55, 56]. Later, active genes located far apart or even on different chromosomes were found

to co-localize at such transcription factories [54]. These same authors showed that two B-cell specific loci, *myc* and *igh*, known as translocation partners in lymphomas, moved into the same preexisting transcription factory upon activation, suggesting that transcription factories may be preexisting structures rather than being the consequence of gene clustering [57]. Others found with minichromosomes containing different transcripts or promoters that only functional related active genes showed co-localization, which led also them to conclude that dedicated factories exist for the transcription of specific subsets of genes [58]. Based on these observations the idea arose that genes need to relocate to preexisting and perhaps even dedicated RNA polII foci for their robust transcription. Visualizing transcription factories in living cells and understanding how they are formed is quite challenging. Recently, a study using single-molecule labeling successfully captured RNA polII clustering in living cells. It showed the dynamic assembling of RNA polII clusters, and measured their average life time as being 5.1 (± 0.4) seconds. This study challenged the concept of pre-existing factories since RNA polII clusters were found to form transiently, with the degree of accumulation strongly being influenced by external stimulation of transcription [59]. The theory of transcription factories driving the co-localization of genes is further challenged by other studies. The inhibition of transcription and removal of RNA PolII from genes did not significantly change their location, demonstrating that genes do not depend on association with transcription factories to maintain their position [60]. Others found no evidence for clustering of co-regulated genes and suggested that they localize at much larger nuclear entities called nuclear speckles, where RNA splicing machinery protein components accumulate [61, 62]. Again it may well be that 'openness' of chromatin, more than transcriptional activity, determines where a gene or a gene locus ends up in the nuclear space [25]. With the development of 3C related methods, spatial contacts can be studied in more details. Also, these technologies enable moving from single locus to genome-wide scale studies. Emerging evidence suggests that master regulators of specific pathways or cell type specific transcription factors may organize the genome and their responsive genes to form subcompartments to facilitate the robust regulation.

Nuclear hormone receptor

A category of proteins that intuitively seems likely to play a role in bringing together enhancers and promoters is the family of nuclear hormone receptors. These proteins migrate to the nucleus upon hormone stimulation to act as transcription factors and activate a large set of target genes. Nuclear hormone receptors work together with chromatin remodelers to manipulate the chromatin structure, in order to achieve a quick response of these target genes. Using an artificial lac operon array it was shown that recruitment of estrogen receptor (ER) is sufficient to unfold this artificial compact chromatin structure, without requirement for transcriptional activation or measurable histone modifications [63]. Another study also reports local chromatin structure changes around estrogen-responsive genes after estrogen treatment [64]. ER may physically remodel chromatin structure to promote a transcriptional response but may also exert its action through changing the 3D genome. Evidence for this is not conclusive though. Microscopy study suggested that androgen receptor (AR) responsive genes move in close proximity and create new cis- and trans-chromosomal

interactions upon AR stimulation [65], but these data could not be confirmed by others [66]. Chromatin Interaction Analysis using Paired-End Tag Sequencing (CHIA-PET), a method that combines chromatin immunoprecipitation (ChIP) with proximity ligation, identified that ER binding sites form extensive interactions upon induction. Genes that react to the induction, mainly those that show up-regulation, often participate in these long-range DNA contacts. Genes which show no effect are outside those ER loops, suggesting that the spatial contacts used by ER are functional related to gene regulation [67]. As CHIA-PET relies on ChIP, it cannot analyze 3D genome structures prior to ER induction. Therefore, whether the observed contacts are formed de novo by ER chromatin binding, or pre-exist to be efficiently used by ER, remains to be determined. At a higher order level genome organization, induction of glucocorticoid receptor (GR)-responsive genes apparently happens without their repositioning or aggregation of target genes in the cell nucleus [68].

Polycomb group proteins

Polycomb group (PcG) proteins are key regulators in chromatin structure, epigenetic regulation, and development. They regulate a large number of genes which are critical for development. In *Drosophila* and mammals, PcG proteins constitute multiple complexes. Among them, polycomb repressive complex1 (PRC1) and polycomb repressive complex2 (PRC2) are the most conserved and characterized ones. In general, PRC2 includes EZH2, which has the ability of depositing H3K27me3, one of the main histone modifications correlated with transcriptional repression. H3K27me3 can attract PRC1, which further stabilizes the binding of the whole complex and induces other posttranslational modifications [69, 70]. The binding sites of PRC1, PRC2 and the distribution of H3K27me3 are well identified in various tissues and cell types. In *Drosophila*, large H3K27me3 domains are identified, which are often associated with inactive genes. In mammalian cells, H3K27me3 constitutes relatively small domains. In ES cells, H3K27me3 domains are not as clearly defined as in differentiated cells. One unique feature of ES cells is that some H3K27me3 sites also carry H3K4me3, a histone modification associated with active transcription. Those unique regions are called "bivalent domains". Most of the genes located in "bivalent domains" are poised for further development [71]. Even though PcG proteins bind to numerous sites in the genome, in *Drosophila* they aggregate in the nucleus as discrete foci, called "PcG bodies" [72, 73]. Genes repressed by PcG proteins often co-localize with "PcG bodies". In *Drosophila*, two genes, located 10MB apart are found to be in one PcG body, when they are both repressed by PcG proteins (Fig.6) [36]. Previous studies suggested that PcG proteins have the ability to compact certain genomic regions. This polycomb-dependent contraction of chromatin is not related to their catalytic features, implying that PcG protein binding is sufficient to directly affect chromatin compaction to achieve efficient regulation of their target genes [74, 75]. By using one of the main targets of PcG as bait in 4C, it was revealed that the targeted locus prefers to interact with similarly typed, PcG-bound chromosomal regions. Depleting one PcG responsive element partially disrupted those extensive contacts and also released the repression of the PcG targeted genes, suggesting the spatial clustering has a functional role in gene regulation [36]. Similar observations were made in mouse ES cells. Genomic regions which are highly enriched for PcG proteins and H3K27me3 show preferential co-

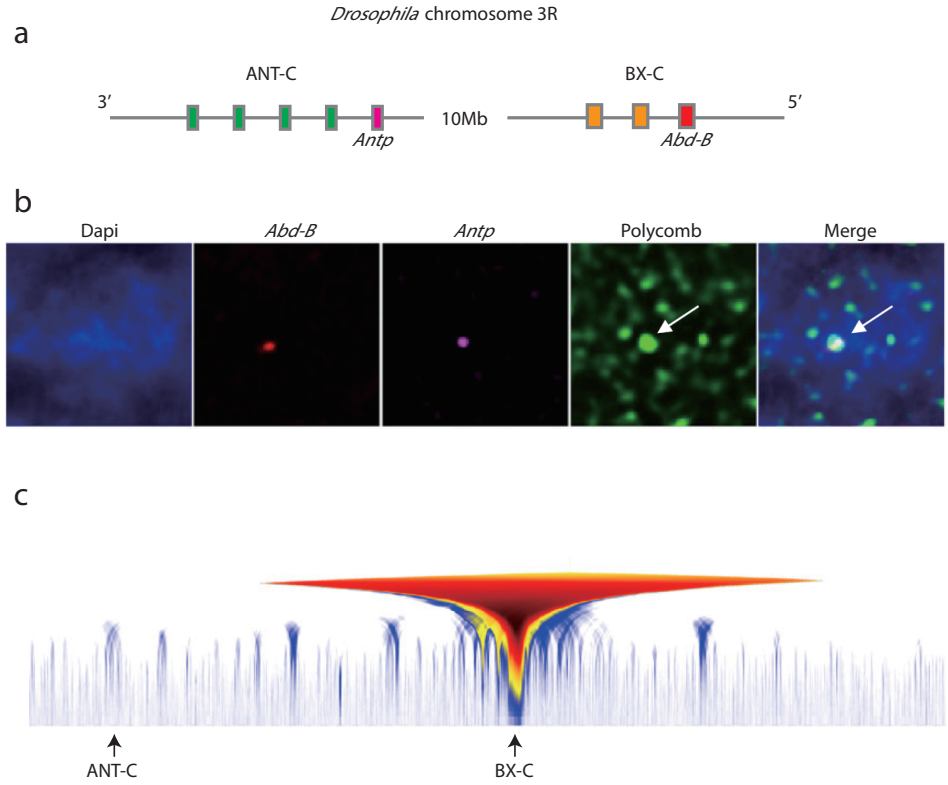


Figure 6. Hox genes kissing in *Drosophila*. In *Drosophila melanogaster*, Hox genes are organized into two complexes: the Antennapedia (ANT-C) and bithorax complex (BX-C). (a) Linear representation of the Hox genes (squares). (b) FISH showing that when gene *Antp* from ANT-C and *Abd-B* from BX-C are both repressed, they co-localize in a "PcG body". (c) 4C domainogram profile using an element from BX-C as a bait, showing BX-C is interacting with ANT-C [36].

localization. These specific contacts could be destroyed by depleting one of the PcG proteins that caused complete loss of H3K27me3; the rest of the global organization remained intact [35]. Those experiments suggest that PcG bodies may serve as silencing compartments that recruit selected, PcG-bound, genes.

CELL TYPE SPECIFIC TRANSCRIPTION FACTOR

Erythroid transcription factors

The topology of the mammalian β-globin locus has been studied extensively by various methods to uncover how the regulatory elements work together in 3D to effect transcription. Using 3C it was shown that the LCR, multiple DHSs spread over nearly 200 kb and the β-globin genes physically come together to form an Active Chromatin Hub (ACH) that

presumably coordinates correct gene expression. 3C analysis has revealed that KLF1, an erythroid-specific transcription factor, is indispensable for establishing such functional ACH [29]. On a genome-wide scale, KLF1 binds to thousands of sites in erythroid cells and many of these bindings site are not close to promoter regions, indicating that long range interactions may commonly be used in KLF1 regulation [76]. A modified 4C study on active mouse globin genes showed that KLF1-regulated genes prefer to cluster at a limited number of foci where KLF1 is concentrated and their specific associations depend on the presence of KLF1 [77]. Those studies suggested that also at a genome-wide scale KLF1 has a spatial function to gather co-regulated genes for their transcription. Perhaps in line with this is the observation that an isolated human β-globin LCR, site-specifically integrated into a mouse chromosome, preferentially aggregated in the nuclear space with target genes of KLF1 and also of GATA1, another erythroid-specific transcription factor that plays a role in chromatin loop formation at the β-globin locus [30]. Both KLF1 and GATA1 bind to multiple sites throughout the LCR. Thus, it seems that genes and regulatory DNA elements controlled by the same transcription factors preferentially meet each other in the nuclear space. This occurs without gross re-shaping of the chromosomes though. The LCR did not get engaged in new long-range contacts as compared to its untargeted integration site but rather selected its preferred partners from pre-existing contacts already formed by this integration-site. Thus, a single enhancer, gene or regulatory DNA element relies mostly on the remainder of its chromosome for its nuclear location [78].

Pluripotency factors

Stem cell research is developing quickly these years. Core pluripotency factors like Nanog, Sox2, Oct4 and Klf2 have been identified, which are essential for keeping the pluripotent state. These factors are also successfully used to reprogram somatic cells back to pluripotent stem cells, known as induced pluripotent stem cells (iPSCs) [79, 80]. Pluripotency factors are well known for working together to control the expression of numerous genes to maintain pluripotency, but whether they play a role in chromatin organization for their regulation is only studied recently. ES cells retain a unique chromatin structure compared with any other differentiated cell type. The chromatin of ES cells is relatively open, structurally less defined, hyperdynamic and permissive for regulators to bind. Pluripotency factors bind to many active enhancer sites [81]. Extensive 4C studies of viewpoints on different pluripotency genes combined with Hi-C data indicated that pluripotency genes and the binding sites of pluripotency factors prefer to cluster with each other in the nucleus. This spatial clustering is only observed in ES cells and iPSCs, and depends on the presence of pluripotency factors. Artificially targeting Nanog to a genomic site unrelated to any pluripotency gene could increase the contact frequency with other Nanog binding sites on the same chromosome, suggesting that Nanog has a direct role in this specific clustering. It is likely that pluripotency factors collaborate with each other to remodel the chromatin structure and the position of genes, which in turn may facilitate the robust control of their gene expression and help keeping the pluripotent status [34, 35]. To extrapolate this: each tissue may have its own specific transcription factors contributing to a unique 3D organization of the genome that best accommodates the required transcriptional outcome.

TETHERING SYSTEMS TO MANIPULATE PROTEIN RECRUITMENT AND GENE POSITION

From the introduction so far it should be clear that over the last decade microscopy and 3C based methods have greatly increased our understanding of the 3D genome and uncovered that genome topology importantly influences genome functioning. Locally, interactions between cis-regulatory elements and promoters are critical for gene regulation. Globally, regulatory factors seem to influence the positioning of their target genes possibly to facilitate their regulation. To better understand the cause and consequence of chromatin loops, gene positioning and transcriptional output, we need to be able to manipulate the structure of the genome. Tethering systems that recruit defined proteins to specific sites in the genome can serve this purpose. Below I will discuss some of the established and upcoming protein recruitment strategies and indicate how they can contribute to a better understanding of the impact of chromosome topology changes (Fig.7).

Lac operon-repressor (LacO-R)/Tet operon-repressor (TetO-R)

The Lac operon-repressor (LacO-R) and Tet operon-repressor (TetO-R) systems are both derived from bacteria and were the first systems used to characterize the consequences of controlled protein recruitment to specific genomic sites. They work in a similar way, taking advantage of the high affinity between the proteins (LacR or TetR) and their cognate DNA binding sequences that can be introduced into the mammalian genome by random insertion or homologous recombination. The proteins can then be fused to selected proteins that may either be fluorescent proteins to follow the loci by live cell imaging, or transcription factors to assay the impact of their recruitment via the efficient binding of LacR to lacO sites. A useful additional feature of the LacO-R system is that LacR binding can be reversed by adding a molecular biology reagent, IPTG [82].

In one of the first studies, large arrays of LacO repeats spanning several tens of mega base pairs (Mb) were used to study the functional consequences of the recruitment of various factors. By targeting for example a strong activator, the acidic activation domain of VP16, to heterochromatinized lacO repeats, large scale de-condensation was observed quickly after targeting, with a strong increase of histone acetylation and co-recruitment of histone acetyltransferases at the targeted site. No transcription was found associated with the initiation and maintenance of this unfolding phenomenon [83]. When targeting HP1 to an extended, euchromatic like array of lacO repeats, local condensation, the accumulation of histone methyltransferase SETB1 and a gain in H3K9me3, a histone modification associated with heterochromatin, was observed [84]. Other factors like estrogen receptor and BRCA1 were studied in the same way to understand how they can affect chromatin [63, 85]. However, the size of these huge arrays raises questions about the physiological relevance of these observations. Shorter versions of lacO repeats (<10 kb) were therefore later used as docking platforms to better study the role of regulatory factor recruitment. KRAB/KAP1, a transcriptional repressor complex, was targeted to various LacO arrays at different genomic locations. Targeting of KRAB/KAP1 appeared to silence endogenous genes neighboring these platforms as far as several tens of kb away. This repressive effect was accompanied

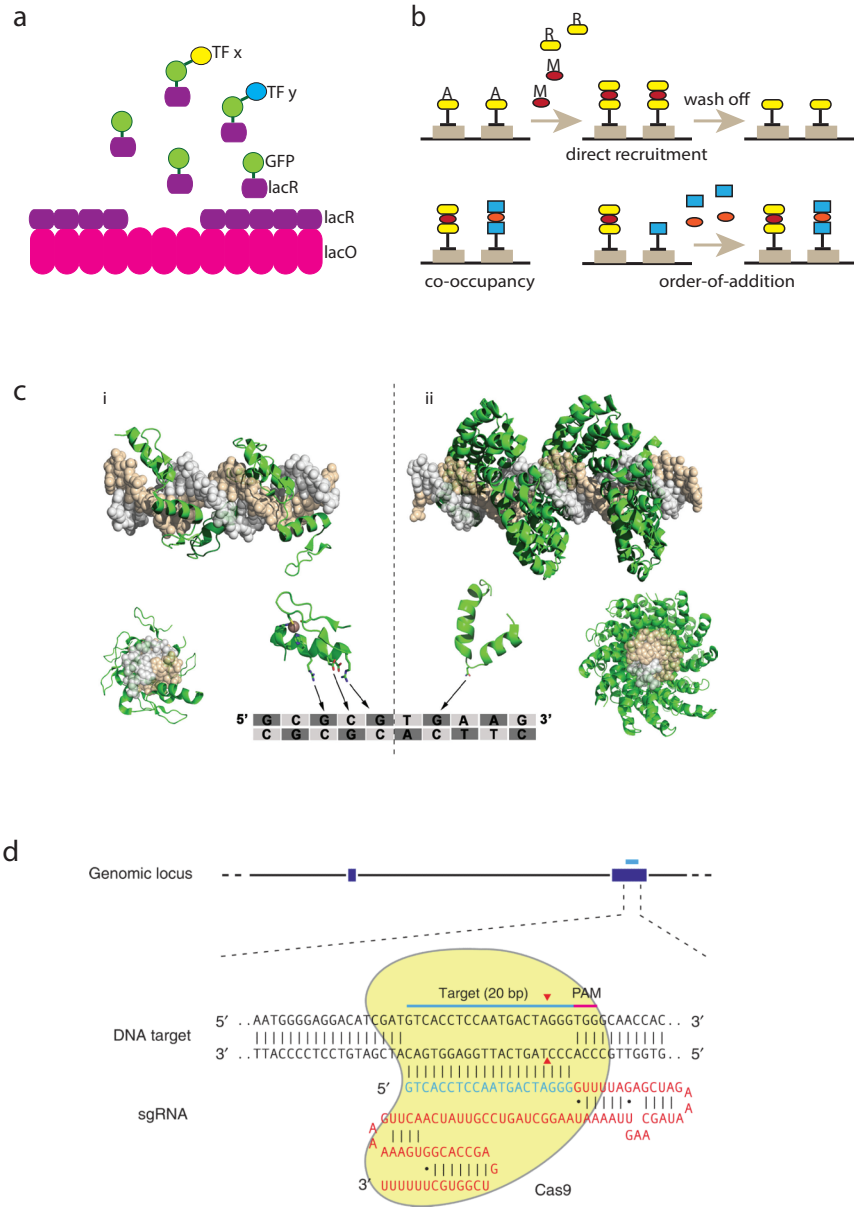


Figure 7. Overview of the tethering systems. (a) Lac operon-repressor assay. TF: transcription factor. (b) Chemical induced proximity assay. A: anchor partner, R: recruitment partner, M: bi-functional molecule. Adapted from [91]. (c) Three dimensional structure of a Zinc Finger Protein and TAL effector. (i) Front and lateral view of a six-finger zinc finger protein. A single zinc finger recognizes 3 bp of DNA. (ii) TAL effectors recognize one single bp of DNA. Adapted from [104]. (d) RNA-guided Cas9 nuclease. The Cas9 nuclease from *S. pyogenes* (in yellow) uses a sgRNA consisting of a 20-nt guide sequence (blue) and a scaffold (red) to target the genomic DNA (blue bar on top strand), directly upstream of a requisite 5′-NGG adjacent motif (PAM; pink). Cas9 induces a DSB ~3 bp upstream of the PAM (red triangle). Adapted from [105].

by the spreading of H3K9me3 and consequently the formation of a new heterochromatin domain. Individual targeting sites showed different levels of reduction of the surrounding genes, which shows that the impact of protein recruitment is position dependent [86].

LacO-R arrays can also be applied to study the functional role of specific nuclear compartments by repositioning genomic regions to these compartments. LacR fused with an inner nuclear membrane protein can drag lacO insertion sites to the nuclear periphery, as was shown for two cassettes inserted in the human genome. After repositioning, some endogenous genes near and a few genes located far (~100MB) from the lacO insertion sites showed down regulation. Many other genes surrounding the lacO arrays retained their normal expression level though. When the LacO-R interaction was abolished by IPTG, the lacO-tagged regions were released from the nuclear periphery, which was accompanied by gene reactivation [87]. These observations indicated that active tethering to the nuclear periphery is sufficient for gene repression which is consistent with the concept that the nuclear periphery is mostly a silent compartment. In another study, the LacO-R array was modified to simultaneously visualize this reposition process and nascent transcription from a co-integrated reporter gene in living cells, during interphase and through mitosis. Here, recruitment to the nuclear periphery was found to only take place after cells going through mitosis. Once recruited, the lacO cassette stably anchored to the nuclear lamina, at least over one cell division. Surprisingly, transcription could be initiated both before and after locus recruitment to the nuclear periphery. Moreover, the kinetics of transcriptional induction was even similar between the targeted and non-targeted transgene [88]. Collectively, this work suggested that the nuclear periphery can induce gene silencing, but that genes respond in a gene specific manner.

The LacO system cannot only be used for recruitment studies, it also enables following the position of individual loci over time in single cells. For example, in such manner X chromosomes pairing has been traced in real time in a study that suggested that pairing is involved in the silencing of one of the X chromosomes in female cells [89]. By combining the LacO-R and the TetO-R system in the same cells one can simultaneously trace the position of two individual loci. This strategy was recently used to follow the fate of translocation partners after digestion with an ectopic restriction site positioned in between the two operon repeats. The study showed that the "contact first" model in double strand break (DSB) repair underlies the vast majority of translocation evens, implying that selection of translocation partners mostly relies on the 3D proximity between genomic regions [90].

Chemical induced proximity (CIP)

Chemical induced proximity (CIP) is yet another system that uses small molecule ligands to achieve targeting at an endogenous locus in vivo. The ligands function by binding two different peptides at the same time. One is known as the CIP anchor, a specific DNA binding domain, which directs where the system is going to land. The other part is the CIP recruiter, fused with a protein of interest. One advantage of using those bi-functional molecules is that they can be easily added and removed from cells, meaning that CIP is a reversible system. Since different pairs of molecules can be combined in CIP, more than one factor can be recruited at the same time or sequentially. Hathaway *et al* developed a chromatin in vivo

assay (CiA) by using CIP in living cells. In ES cells, one of the Oct4 alleles was modified to have two CIP anchor sites. HP1a was first targeted to active Oct4 locus by CiA. Gene silencing and H3K9me3 accumulation were immediately induced after targeting. H3K9me3 could spread 10kb away from the targeting site, forming small heterochromatic domain. After removing the binding of HP1a, H3K9me3 stably transmitted over few generations, but the repression of gene was partially passed off. In contrast, when targeting a strong activator VP16 to inactive Oct4 locus in mouse embryonic fibroblasts (MEFs), Oct4 got reactivated afterwards. This reactivation could be reversed by retargeting HP1a to the same site. As soon as HP1a binding there, the repression and negative histone modification were quickly reestablished. However, in this situation, H3K9me3 did not spread out as far as in ES cell. It only accumulated at the Oct4 promoter region [91]. The different targeting results in those two cell lines imply that even though transcription factors have a strong influence on a single locus, as regulating expression and inducing histone modification, their ability is limited by the local chromatin environment.

Zinc Finger (ZF) and Transcription activator-like effector (TALE)

Zinc Fingers (ZFs) are small protein structural motifs. They can bind DNA, RNA and other molecules. Engineered tandem repeats of ZFs can bind specific DNA sequences. Transcription activator-like effectors (TALEs) are DNA-binding proteins from *Xanthomonas* bacteria, which consist of tandem repeats of 34-amino acid repeat that can be engineered to target DNA sequence. In ZF array, each module targets three DNA bases. When combining different ZF modules, the binding capability always needs to be tested first. The monomers from TALE can recognize single nucleotide, which makes TALEs more flexible for designing purposes. Both ZFs and TALEs can be coupled with other protein domains and as such serve as useful genome engineering tools. Initially, ZFs and TALEs were used to alter the expression of selected genes by targeting transcriptional activator or repressor domains to specific upstream sequences [92-95]. Recently, the strategies have also been applied to perturb the local chromatin conformation and study the functional consequences. In GATA1 null erythroid cell, the β-globin locus is inactive and there is no physical contact between LCR and the distal β-globin gene promoter. By ZF-mediated targeting Ldb1, a factor belonging to various erythroid transcription factor complexes, to the β-globin gene promoter, a loop was formed between the LCR and the β-globin promoter. This physical interaction partially rescues β-globin gene expression, indicating that forced loop formation between an enhancer and promoter is sufficient for transcription [96]. ZF and TALE combined with nucleases can also introduce DSB at specific genomic sites, which can also be used as an efficient and precise way to cleave loops [97]. The application of these new tool boxes opens new avenues for studying chromosomal contacts and their functional consequences on gene transcription.

CRISPR-Cas9

CRISPR-Cas9 was first developed as a genome engineering system. It uses a single-guide RNA (sgRNA) containing 20 nucleotides homology to the target as a guide sequence, and Cas9 to then locally induce a DSB. A modified version of CRISPR-Cas9 has been applied as a tethering system to activate and repress specific genes in bacterial, yeast and human

cells. In this system, Cas9 was mutated to lose its nuclease ability and instead was fused with a protein of interest [98]. Due to the easy and flexible design of sgRNA, and the fact that different sgRNAs can be used in combination with different Cas proteins, it opens the possibility to simultaneously target different factors to different endogenous DNA sequences [99]. As such, also this system seems highly suitable to create artificial chromatin loops in the genome and study how proximity between distal DNA sequences influences processes like transcription.

In this thesis, chapter 2 describes that we applied lacO-R array to systematically study the role of transcription factors in mediating chromatin interactions and the functional consequences of gene repositioning by targeting different transcription factors to the lacO insertion sites. In chapter 3, we studied how pluripotency factors contribute to the 3D organization of the genome by combing 4C and Hi-C data. The direct role of pluripotency factor Nanog was studied by targeting Nanog to a genomic site through lacO-R array. Additionally, the structural role of histone H1 was investigated in chapter 4, by comparing Hi-C profiles and other genome-wide profiles between ES cells with different levels of histone H1.

REFERENCES

1. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
2. *The ENCODE (ENCyclopedia Of DNA Elements) Project.* Science, 2004. **306**(5696): p. 636-40.
3. Crawford, G.E., et al., *DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays.* Nat Methods, 2006. **3**(7): p. 503-9.
4. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.
5. Consortium, E.P., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
6. Almouzni, G. and A.V. Probst, *Heterochromatin maintenance and establishment: lessons from the mouse pericentromere.* Nucleus, 2011. **2**(5): p. 332-8.
7. Woodcock, C.L. and R.P. Ghosh, *Chromatin higher-order structure and dynamics.* Cold Spring Harb Perspect Biol, 2010. **2**(5): p. a000596.
8. Cremer, M., et al., *Inheritance of gene density-related higher order chromatin arrangements in normal and tumor cell nuclei.* J Cell Biol, 2003. **162**(5): p. 809-20.
9. Croft, J.A., et al., *Differences in the localization and morphology of chromosomes in the human nucleus.* J Cell Biol, 1999. **145**(6): p. 1119-31.
10. Goetze, S., et al., *The three-dimensional structure of human interphase chromosomes is related to the transcriptome map.* Mol Cell Biol, 2007. **27**(12): p. 4475-87.
11. Bolzer, A., et al., *Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes.* PLoS Biol, 2005. **3**(5): p. e157.
12. Branco, M.R., et al., *Changes in chromosome organization during PHA-activation of resting human lymphocytes measured by cryo-FISH.* Chromosome Res, 2008. **16**(3): p. 413-26.
13. Arsuaga, J., et al., *Chromosome spatial clustering inferred from radiogenic aberrations.* Int J Radiat Biol, 2004. **80**(7): p. 507-15.
14. Volpi, E.V., et al., *Large-scale chromatin organization of the major histocompatibility complex and other regions of human chromosome 6 and its response to interferon in interphase nuclei.* J Cell Sci, 2000. **113 (Pt 9)**: p. 1565-76.

15. Williams, R.R., et al., *Subchromosomal positioning of the epidermal differentiation complex (EDC) in keratinocyte and lymphoblast interphase nuclei.* Exp Cell Res, 2002. **272**(2): p. 163-75.
16. Mahy, N.L., P.E. Perry, and W.A. Bickmore, *Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH.* J Cell Biol, 2002. **159**(5): p. 753-63.
17. Ferrai, C., et al., *Poised transcription factories prime silent uPA gene prior to activation.* PLoS Biol, 2010. **8**(1): p. e1000270.
18. Brown, K.E., et al., *Dynamic repositioning of genes in the nucleus of lymphocytes preparing for cell division.* Mol Cell, 1999. **3**(2): p. 207-17.
19. Kosak, S.T., et al., *Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development.* Science, 2002. **296**(5565): p. 158-62.
20. Williams, R.R., et al., *Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus.* J Cell Sci, 2006. **119**(Pt 1): p. 132-40.
21. Zink, D., et al., *Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei.* J Cell Biol, 2004. **166**(6): p. 815-25.
22. Skok, J.A., et al., *Nonequivalent nuclear location of immunoglobulin alleles in B lymphocytes.* Nat Immunol, 2001. **2**(9): p. 848-54.
23. Holwerda, S.J., et al., *Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells.* Nucleic Acids Res, 2013. **41**(14): p. 6905-16.
24. Morey, C., et al., *Nuclear reorganisation and chromatin decondensation are conserved, but distinct, mechanisms linked to Hox gene activation.* Development, 2007. **134**(5): p. 909-19.
25. Chambeyron, S. and W.A. Bickmore, *Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription.* Genes Dev, 2004. **18**(10): p. 1119-30.
26. Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. **295**(5558): p. 1306-11.
27. Tolhuis, B., et al., *Looping and interaction between hypersensitive sites in the active beta-globin locus.* Mol Cell, 2002. **10**(6): p. 1453-65.
28. Palstra, R.J., et al., *The beta-globin nuclear compartment in development and erythroid differentiation.* Nat Genet, 2003. **35**(2): p. 190-4.
29. Drissen, R., et al., *The active spatial organization of the beta-globin locus requires the transcription factor EKLF.* Genes Dev, 2004. **18**(20): p. 2485-90.
30. Vakoc, C.R., et al., *Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1.* Mol Cell, 2005. **17**(3): p. 453-62.
31. Splinter, E., et al., *CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus.* Genes Dev, 2006. **20**(17): p. 2349-54.
32. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* Nat Genet, 2006. **38**(11): p. 1348-54.
33. Splinter, E., et al., *The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA.* Genes Dev, 2011. **25**(13): p. 1371-83.
34. de Wit, E., et al., *The pluripotent genome in three dimensions is shaped around pluripotency factors.* Nature, 2013. **501**(7466): p. 227-31.
35. Denholtz, M., et al., *Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization.* Cell Stem Cell, 2013. **13**(5): p. 602-16.
36. Bantignies, F., et al., *Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila.* Cell, 2011. **144**(2): p. 214-26.
37. Sanyal, A., et al., *The long-range interaction landscape of gene promoters.* Nature, 2012. **489**(7414): p. 109-13.
38. van de Werken, H.J., et al., *Robust 4C-seq data analysis to screen for regulatory DNA interactions.* Nat Methods, 2012. **9**(10): p. 969-72.

39. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.
40. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.
41. Nora, E.P., et al., *Spatial partitioning of the regulatory landscape of the X-inactivation centre.* Nature, 2012. **485**(7398): p. 381-5.
42. Nagano, T., et al., *Single-cell Hi-C reveals cell-to-cell variability in chromosome structure.* Nature, 2013. **502**(7469): p. 59-64.
43. Dechat, T., et al., *Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin.* Genes Dev, 2008. **22**(7): p. 832-53.
44. Vergnes, L., et al., *Lamin B1 is required for mouse development and nuclear integrity.* Proc Natl Acad Sci U S A, 2004. **101**(28): p. 10428-33.
45. Sullivan, T., et al., *Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy.* J Cell Biol, 1999. **147**(5): p. 913-20.
46. Hale, J.S., et al., *Cell-extrinsic defective lymphocyte development in Lmna(-/-) mice.* PLoS One, 2010. **5**(4): p. e10127.
47. Malhas, A., et al., *Defects in lamin B1 expression or processing affect interphase chromosome position and gene expression.* J Cell Biol, 2007. **176**(5): p. 593-603.
48. Guelen, L., et al., *Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.* Nature, 2008. **453**(7197): p. 948-51.
49. Peric-Hupkes, D., et al., *Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation.* Mol Cell, 2010. **38**(4): p. 603-13.
50. Pickersgill, H., et al., *Characterization of the Drosophila melanogaster genome at the nuclear lamina.* Nat Genet, 2006. **38**(9): p. 1005-14.
51. Zullo, J.M., et al., *DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina.* Cell, 2012. **149**(7): p. 1474-87.
52. Meuleman, W., et al., *Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence.* Genome Res, 2013. **23**(2): p. 270-80.
53. Kind, J., et al., *Single-cell dynamics of genome-nuclear lamina interactions.* Cell, 2013. **153**(1): p. 178-92.
54. Osborne, C.S., et al., *Active genes dynamically colocalize to shared sites of ongoing transcription.* Nat Genet, 2004. **36**(10): p. 1065-71.
55. Pombo, A., et al., *Specialized transcription factories within mammalian nuclei.* Crit Rev Eukaryot Gene Expr, 2000. **10**(1): p. 21-9.
56. Ferrai, C., et al., *Gene positioning.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000588.
57. Osborne, C.S., et al., *Myc dynamically and preferentially relocates to a transcription factory occupied by Igh.* PLoS Biol, 2007. **5**(8): p. e192.
58. Xu, M. and P.R. Cook, *Similar active genes cluster in specialized transcription factories.* J Cell Biol, 2008. **181**(4): p. 615-23.
59. Cisse, II, et al., *Real-time dynamics of RNA polymerase II clustering in live human cells.* Science, 2013. **341**(6146): p. 664-7.
60. Palstra, R.J., et al., *Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription.* PLoS One, 2008. **3**(2): p. e1661.
61. Moen, P.T., Jr., et al., *Repositioning of muscle-specific genes relative to the periphery of SC-35 domains during skeletal myogenesis.* Mol Biol Cell, 2004. **15**(1): p. 197-206.
62. Brown, J.M., et al., *Association between active genes occurs at nuclear speckles and is modulated by chromatin environment.* J Cell Biol, 2008. **182**(6): p. 1083-97.
63. Nye, A.C., et al., *Alteration of large-scale chromatin structure by estrogen receptor.* Mol Cell Biol, 2002. **22**(10): p. 3437-49.

64. Osmanbeyoglu, H.U., et al., *Estrogen represses gene expression through reconfiguring chromatin structures.* Nucleic Acids Res, 2013. **41**(17): p. 8061-71.
65. Lin, C., et al., *Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer.* Cell, 2009. **139**(6): p. 1069-83.
66. Kocanova, S., et al., *Activation of estrogen-responsive genes does not require their nuclear co-localization.* PLoS Genet, 2010. **6**(4): p. e1000922.
67. Fullwood, M.J., et al., *An oestrogen-receptor-alpha-bound human chromatin interactome.* Nature, 2009. **462**(7269): p. 58-64.
68. Hakim, O., et al., *Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements.* Genome Res, 2011. **21**(5): p. 697-706.
69. Schuettengruber, B., et al., *Genome regulation by polycomb and trithorax proteins.* Cell, 2007. **128**(4): p. 735-45.
70. Sparmann, A. and M. van Lohuizen, *Polycomb silencers control cell fate, development and cancer.* Nat Rev Cancer, 2006. **6**(11): p. 846-56.
71. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
72. Saurin, A.J., et al., *The human polycomb group complex associates with pericentromeric heterochromatin to form a novel nuclear domain.* J Cell Biol, 1998. **142**(4): p. 887-98.
73. Orlando, V., *Polycomb, epigenomes, and control of cell identity.* Cell, 2003. **112**(5): p. 599-606.
74. Eskeland, R., et al., *Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination.* Mol Cell, 2010. **38**(3): p. 452-64.
75. Margueron, R., et al., *Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms.* Mol Cell, 2008. **32**(4): p. 503-18.
76. Tallack, M.R., et al., *Novel roles for KLF1 in erythropoiesis revealed by mRNA-seq.* Genome Res, 2012. **22**(12): p. 2385-98.
77. Schoenfelder, S., et al., *Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells.* Nat Genet, 2010. **42**(1): p. 53-61.
78. Noordermeer, D., et al., *Variegated gene expression caused by cell-specific long-range DNA interactions.* Nat Cell Biol, 2011. **13**(8): p. 944-51.
79. Takahashi, K., et al., *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* Cell, 2007. **131**(5): p. 861-72.
80. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* Cell, 2006. **126**(4): p. 663-76.
81. Marson, A., et al., *Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.* Cell, 2008. **134**(3): p. 521-33.
82. Cronin, C.A., W. Gluba, and H. Scrable, *The lac operator-repressor system is functional in the mouse.* Genes Dev, 2001. **15**(12): p. 1506-17.
83. Tumbar, T., G. Sudlow, and A.S. Belmont, *Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain.* J Cell Biol, 1999. **145**(7): p. 1341-54.
84. Verschure, P.J., et al., *In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation.* Mol Cell Biol, 2005. **25**(11): p. 4552-64.
85. Ye, Q., et al., *BRCA1-induced large-scale chromatin unfolding and allele-specific effects of cancer-predisposing mutations.* J Cell Biol, 2001. **155**(6): p. 911-21.
86. Groner, A.C., et al., *KRAB-zinc finger proteins and KAP1 can mediate long-range transcriptional repression through heterochromatin spreading.* PLoS Genet, 2010. **6**(3): p. e1000869.
87. Finlan, L.E., et al., *Recruitment to the nuclear periphery can alter expression of genes in human cells.* PLoS Genet, 2008. **4**(3): p. e1000039.
88. Kumaran, R.I. and D.L. Spector, *A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence.* J Cell Biol, 2008. **180**(1): p. 51-65.

89. Masui, O., et al., *Live-cell chromosome dynamics and outcome of X chromosome pairing events during ES cell differentiation.* Cell, 2011. **145**(3): p. 447-58.
90. Roukos, V., et al., *Spatial dynamics of chromosome translocations in living cells.* Science, 2013. **341**(6146): p. 660-4.
91. Hathaway, N.A., et al., *Dynamics and memory of heterochromatin in living cells.* Cell, 2012. **149**(7): p. 1447-60.
92. Beerli, R.R., B. Dreier, and C.F. Barbas, 3rd, *Positive and negative regulation of endogenous genes by designed transcription factors.* Proc Natl Acad Sci U S A, 2000. **97**(4): p. 1495-500.
93. Gommans, W.M., H.J. Haisma, and M.G. Rots, *Engineering zinc finger protein transcription factors: the therapeutic relevance of switching endogenous gene expression on or off at command.* J Mol Biol, 2005. **354**(3): p. 507-19.
94. Zhang, F., et al., *Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription.* Nat Biotechnol, 2011. **29**(2): p. 149-53.
95. Klug, A., *The discovery of zinc fingers and their development for practical applications in gene regulation and genome manipulation.* Q Rev Biophys, 2010. **43**(1): p. 1-21.
96. Deng, W., et al., *Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor.* Cell, 2012. **149**(6): p. 1233-44.
97. Fanucchi, S., et al., *Chromosomal contact permits transcription between coregulated genes.* Cell, 2013. **155**(3): p. 606-20.
98. Gilbert, L.A., et al., *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* Cell, 2013. **154**(2): p. 442-51.
99. Qi, L.S., et al., *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
100. Olins, D.E. and A.L. Olins, *Granulocyte heterochromatin: defining the epigenome.* BMC Cell Biol, 2005. **6**: p. 39.
101. Gheldof, N., et al., *Cell-type-specific long-range looping interactions identify distant regulatory elements of the CFTR gene.* Nucleic Acids Res, 2010. **38**(13): p. 4325-36.
102. Zhang, Y., et al., *Spatial organization of the mouse genome and its role in recurrent chromosomal translocations.* Cell, 2012. **148**(5): p. 908-21.
103. Dekker, J., M.A. Marti-Renom, and L.A. Mirny, *Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data.* Nat Rev Genet, 2013. **14**(6): p. 390-403.
104. Perez-Pinera, P., D.G. Ousterout, and C.A. Gersbach, *Advances in targeted genome editing.* Curr Opin Chem Biol, 2012. **16**(3-4): p. 268-77.
105. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-308.

Transcription-independent control of higher-order chromosome topology by trans-acting factors

Patrick J.Wijchers[1]#, Peter H. L. Krijger[1]#, Yun Zhu[1]#, Diewertje G. E. Piebes[2], Pernette J. Verschure[2], Marjon J. A. M. Verstegen[1], Mark Janssen[1], Hans Teunissen[1], Geert Geeven[1] & Wouter de laat[1]

# Equal contribution

[1] Hubrecht Institute-KNAW & University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

[2] Synthetic Systems Biology and Nuclear Organization Group, Swammerdam Institute for Life Sciences, University of Amsterdam, 1098 XH Amsterdam, The Netherlands.

*Correspondence should be addressed to W.d.L. (w.delaat@hubrecht.eu)

MANUSCRIPT IN PREPARATION

ABSTRACT

Microscopy studies and high throughput chromosome conformation capture (3C) techniques have shown that different types of chromatin are kept separate in the nucleus. Active and inactive chromosomal segments occupy distinct locations relative to nuclear landmark structures, suggesting that the nuclear organization of chromosomes is associated with regulation of gene expression. However, many questions remain concerning the flexibility of chromosomal regions to change nuclear position and the influence of transcription and chromatin status.

To investigate the mechanisms that drive the spatial arrangement and position of chromosomes in the nucleus, we recruited different types of repressive chromatin regulators to lacO/lacR recruitment platforms integrated in transcriptionally active genomic loci in the same cell. We found that genomic loci differ in their propensity to take up new positions in the nucleus. Spatial repositioning was not driven by gene expression changes, but by the associated trans-acting factors. Although all repressive, each chromatin protein redirected a locus specifically towards areas enriched for the same type of chromatin that it deposits. This movement was not limited to the lacO array itself, but involved the whole sub-megabase scaled topologically associating domain. Finally, spatial repositioning did not affect gene expression. Altogether, our data suggests that gene expression and higher-order chromosome topology beyond that of topologically associating domains are not causally related, but independently controlled by the locally associated trans-acting factors.

INTRODUCTION

Nuclear organization of interphase chromosomes is considered to be an important regulator of genome function. Chromosomes occupy non-random positions in the nucleus, and the position of a gene relative to nuclear compartments is thought to influence the probability of expression [1]. Microscopy studies and high throughput chromosome conformation capture (3C) techniques [2, 3] have greatly increased our knowledge of the three-dimensional configuration of the genome in the nucleus[4]. However, the mechanisms that govern the spatial arrangement of chromosomes and gene positioning in the nucleus are still poorly understood.

Individual chromosomes occupy visually discrete volumes within the nucleus ('chromosome territories')[5], and genome-wide interaction profiles have shown that genomic elements preferentially contact regions on the same chromosome[6, 7]. Hi-C and 5C studies revealed that chromosomes are partitioned into hundreds of distinct chromatin interaction domains within which local contacts occur much more frequently than between domains [8, 9]. These megabase-scaled 'topologically associating domains' (TADs) are highly conserved across tissues, and most physical contacts between a gene promoter and its distal regulatory elements will take place within such domains [10, 11]. TADs vary in their transcriptional activity, DNA sequence and chromatin composition [12][7], and TADs preferentially associate with domains of comparable chromatin type [7, 13]. In Hi-C data, this segregation of different chromatin types in the nucleus is apparent as a biphasic state with an A and B compartment [7], which have recently been subdivided into A1-2 and B1-4, each with their own chromatin signature [14]. The A compartment represents colocalization of transcriptionally active or permissive chromatin, and B the spatial crowding of inactive chromatin at the nuclear periphery [14, 15], nucleoli [14, 16, 17] and chromocenters (Wijchers, Geeven et al., submitted). In contrast to TADs, whose borders are mostly conserved across cell types [8], these A and B compartments are linked to gene expression status and therefore cell type-specific. Indeed, in a recent comparison of 5 different human lineages approximately one third of the genome switched A/B compartments in at least one lineage [11]. Although this strong correlation suggests a role for nuclear compartmentalization in gene regulation, to what extent these cell type-specific compartment switches are a cause or consequence of gene activity is largely unclear.

Cell type-specific nuclear arrangements imply that genomic loci must possess an intrinsic flexibility to allow repositioning in response to differentiation signals. Indeed there are examples of endogenous genes that appear to change nuclear positions during lineage commitment, as judged from fluorescence in situ hybridization (FISH) studies. When silenced, they tend to locate to more peripheral nuclear locations in the nucleus [18, 19] or closer to chromocenters [20], whereas they adopt a more internal nuclear position when activated [19, 21]. Genes are physical parts of much larger chromosomes and will therefore rely on neighbouring chromosomal segments for their freedom to move. Neuronal differentiation-induced activation of Mash1 is accompanied by relocation towards the nuclear centre, which drags along immediately flanking genes that nevertheless remain transcriptionally silent [22]. Vice versa, the ectopic integration of sequences that

induce recruitment to the nuclear lamina was found to lead to corresponding peripheral repositioning which in this case was accompanied by the silencing of surrounding genes [23]. Thus, peripheral positioning correlates with, but does not strictly imply, gene silencing. Live cell microscopy studies confirmed this and demonstrated that gene dynamics is generally constrained during interphase [24]. These studies showed that, with a few exceptions [25, 26], gene repositioning requires mitosis, with new locations only being adopted in daughter cells [27-30].

So far, studies designed to experimentally alter the protein composition of a given gene locus and follow its nuclear positioning in a given cell type relied on microscopic techniques that measure location relative to visibly distinct nuclear structures such as the nuclear periphery, chromocenters, nucleoli, polycomb bodies, speckles, transcription factories or other gene loci [31, 32]. These nuclear landmarks all contain different DNA- and chromatin-binding proteins. The general perception from these studies is that tethering of compartment-specific proteins leads to locus recruitment to the corresponding nuclear entity [27, 28, 30, 33]. How to reconcile this with the biphasic genome conformation appreciated from Hi-C studies? Few studies used 3C-based methods to follow conformational changes upon protein tethering to a locus. 4C technology is highly suited for this, as it unbiasedly assays the genome-wide contacts of a locus of choice [6]. Blocking transcription and depleting RNA polymerase II from active gene loci did not cause any major changes in their genome-wide 4C contact maps [34], nor did the recruitment of the glucocorticoid receptor to its target genes [35]. The introduction of a 'super enhancer' (the beta-globin locus control region, which recruits a known set of transcription factors including GATA1 and Klf1) into an already active chromosomal region did not induce new intra- and interchromosomal contacts, as judged from 4C, but specifically strengthened those already formed with regions rich in binding sites for the same transcription factors [36]. A similarly subtle change in preferred contacts with cognate chromosomal regions was seen when the pluripotency factor Nanog was recruited to an artificial lac operator cassette [37]. These latter observations suggest that genes can reside in more than just two (A and B) genomic environments, a notion that receives support from data suggesting that various types of chromosomal regions with similar chromatin signatures preferentially cluster in the nuclear space [14, 38-41]. However, these findings are generally based on genome-wide correlations, but little evidence from experimental locus repositioning is available to support these concepts.

To start exploring the flexibility of different chromosomal regions to migrate to various genomic environments, to assess how this relates to transcriptional changes and epigenetic features, and to ask how much of surrounding sequences is moving along, we have recruited different types of repressive chromatin regulators to lacO/lacR recruitment platforms integrated in transcriptionally active genomic loci in the same cell. We found that the susceptibility to spatial repositioning depends on genomic context and that the degree and direction of movement is determined by the associated trans-acting factor. Locus repositioning is uncoupled from transcriptional changes. Rather, the different local chromatin signatures appear to influence the direction of movement in the nucleus, with each repressive protein inducing a unique shift towards regions that naturally associate with

this factor. Repositioning is not limited to the recruitment platform, but involves all genes within the whole (sub) TAD and influencing also the location of directly neighboring TADs. Our data suggest that gene expression and higher-order chromosome topology beyond that of TADs are not causally related but independently controlled by the locally associated trans-acting factors.

RESULTS

Experimental manipulation of local chromatin conformation

To manipulate the local chromatin conformation at defined genomic loci, we sequentially integrated a recruitment platform into two genomic regions of the same F1 ES cells (ESCs) derived from C57Bl/6 and 129/Sv mouse strains (**Fig. 1A**). The recruitment platform consisted of a kanamycin resistance gene flanked on either side by an array of 120 lac operator (lacO) repeats [42], that serve as high affinity binding sites for the *Escherichia coli* Lac repressor (lacR) (**Fig. 1A**). After integration of the first array on chromosome 8 (previously used in [37]), the additional floxed neomycin resistance gene was removed to allow a second integration on chromosome 11 in the same cells. Correct and single integrations were validated by Southern blot (**Fig. S1**).

We then expressed 3 different transcription and chromatin factors fused to an enhanced green fluorescent protein (EGFP)-lacR moiety to direct these proteins to the lacO cassettes. These factors included Ezh2, the component of the repressive polycomb group protein complex PRC2 that is responsible for H3K27 trimethylation, Suv39h1, the constitutive heterochromatin histone methyltransferase factor that deposits H3K9me3 and Nanog, the pluripotency transcription factor that can have an activating as well as repressive effect on transcription (REF). Viral transduction led to expression of fusion proteins of the expected size (**Fig. 1B**). EGFP-lacR showed a uniform nuclear distribution, except for two bright spots marking the lacO arrays (**Fig. 1C**). These two bright foci were also visible in EGFP-lacR-Nanog cells, which otherwise showed a more grainy pattern than EGFP-lacR alone, consistent with the thousands of Nanog binding sites across the genome [43]. EGFP-lacR-Ezh2 was found throughout the nucleus with multiple bright foci, possibly highlighting the presence of Polycomb bodies [44], whereas EGFP-lacR-Suv39h1 showed the typical enrichment in DAPI-dense pericentromeric heterochromatin foci [45, 46]. Binding of the fusion proteins to the lacO arrays was further verified by ChIP using antibodies against GFP (**Fig. 1D**). Thus, all fusion proteins behaved as expected and bound to the lacO arrays.

Susceptibility to spatial repositioning depends on genomic location and trans-acting factors, but not on transcriptional changes

To explore the ability of the selected regulatory proteins to change the nuclear position of the lacO loci, we applied 4C-seq [47] to probe for chromosomal regions that frequently contact the lacO loci. For the lacO locus on chromosome 11 (chr11), we used a viewpoint in the neomycin resistance (*neo*) gene. Since the *neo* gene was deleted from the lacO cassette

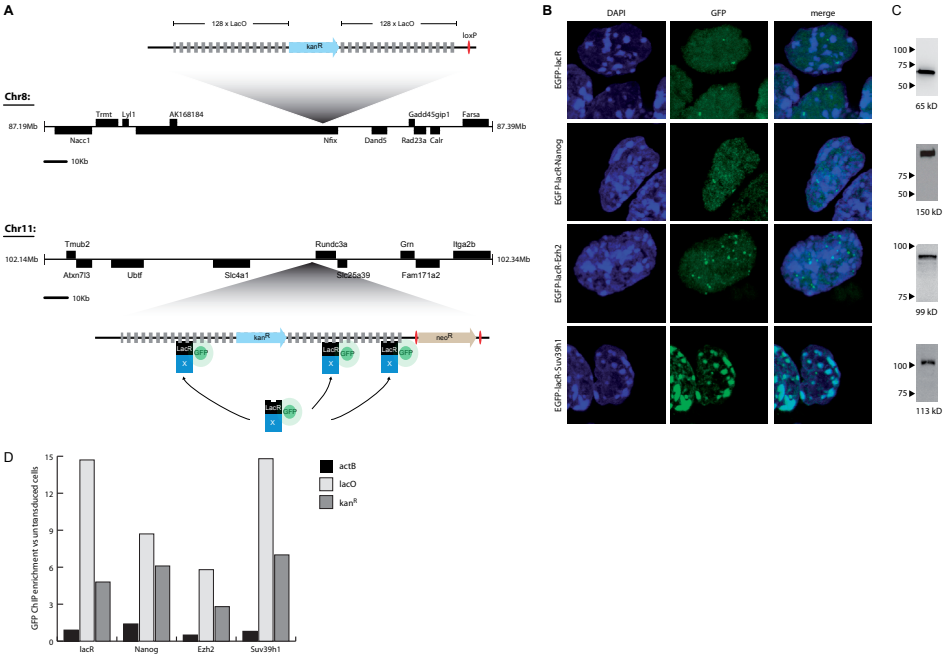


Figure 1. A lacO/lacR recruitment platform to induce nuclear repositioning

A. Schematic view of genomic environment of laco integration sites on chr8 and chr11. Trans-acting factors are recruited to the lacO array when fused to EGFP-lacR.

B. Nuclear distributions of the fusion proteins used. Left panel shows DAPI-stained nuclei, middle panel displays GFP signal, with an overlay in the third panel.

C. Western blots confirming fusion proteins are of the correct size. Numbers indicate size of protein markers (left) and expected size of fusion protein (below).

ChIP with GFP antibody to confirm binding of each protein to the lacO cassette. Input-normalized values are expressed as enrichment over untransduced cells.

on chromosome 8 (chr8), we used a viewpoint 30 kb upstream of the lacO array where the read primer was designed to identify a SNP between the 129 and Bl6 alleles. This allowed for direct comparison of contact profiles between the untargeted and the lacO-transgenic allele. As expected from Hi-C data which shows that chromosomal sequences position themselves as large domains, the overall chromosome and genome-wide contact profiles of viewpoints 30kb apart are highly similar, as confirmed by analysis of a neo viewpoint on chromosome 8 in the single targeted cells (**Fig. S2**). The viewpoint at 30kb from the LacO cassette can therefore serve to follow its positioning. Chromosome-wide 4C profiles were analysed using domainograms [48] to visualise regions that showed statistically significant interactions at different scales (**Fig. 2**).

We found that binding of each chromatin proteins had little impact on the overall chromosome topology of the lacO locus on chr8 (**Fig. 2A**). This was consistent with our

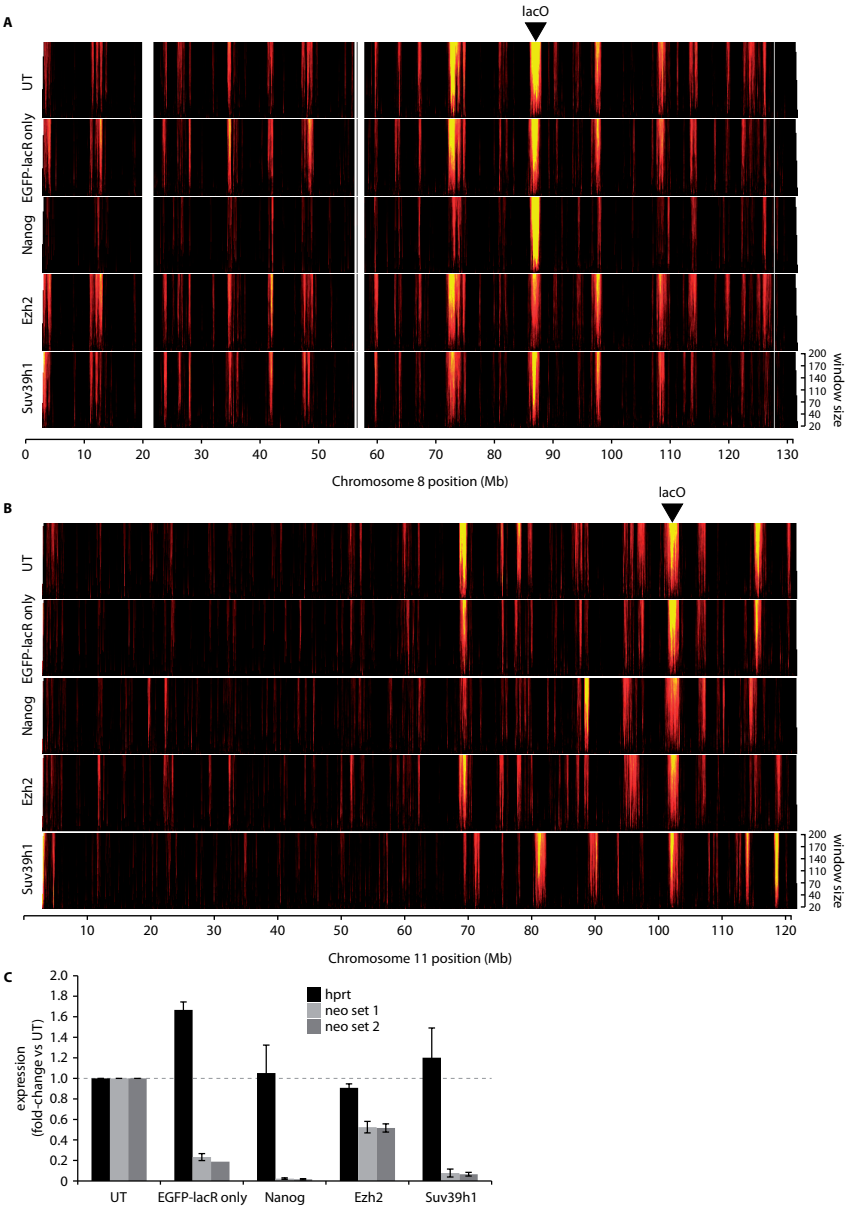


Figure 2. Susceptibility to spatial repositioning depends on genomic location and trans-acting factors, but not on transcriptional changes

A. Domainograms showing allele-specific 4C profiles for the BI/6 allele from the chr8 viewpoint in untransduced (UT) and transduced ES cells. The position of the lacO array is indicated. Names of fusion proteins are abbreviated to just Nanog, Ezh2 and Suv39h1.

B. As in A), but looking from the neo gene in the lacO cassette on chr11.

C. Quantitative RT-PCR analysis of neomycin (2 primer sets) and control genes. Data was normalized to actB and expressed as fold-change over untransduced (UT) cells.

previous finding that Nanog only induced small scale quantitative changes upon binding to the lacO array on chr8 [37]. Conversely, in these same cells all three chromatin proteins influenced the contact profiles of the lacO locus on chr11, albeit to different extents (**Fig. 2B**). These changes were specific for each protein, as EGFP-lacR alone showed little or no changes compared to untransduced (UT) cells. Nanog and Ezh2 induced a small number of new interactions (**Fig. 2B**), but the otherwise conserved contact profiles suggests that only a small proportion of cells engaged in new connections with other genomic regions. Strikingly, Suv39h1 binding appeared to completely rewire the chromosome topology of the chr11 lacO locus, with most common interactions lost and replaced by new ones. Thus, genomic loci differ markedly in their susceptibility to change nuclear position in response to binding of chromatin proteins.

As nuclear position is often linked to gene expression, we wondered whether transcriptional changes could be a driving force on the repositioning of the lacO locus. As expected from their known roles in repressive complexes [49-53], expression levels of the *neo* gene in the chr11 lacO transgene was decreased by binding of Ezh2 and Suv39h1 (**Fig. 2C**). The same was true for Nanog, which also had a strong repressive effect on *neo* transcription (**Fig. 2C**). In fact, EGFP-lacR by itself already acted as a repressor on this locus, which has been attributed to the tight binding of lacR to DNA [54]. However, repression by EGFP-lacR did not lead to changes in chromosome conformation of the lacO locus on chr11, implying that spatial repositioning of this locus is not driven by transcriptional changes. Altogether, we conclude that susceptibility of a locus to change its spatial conformation depends on genomic location and trans-acting factor association, but not on transcriptional status.

Local chromatin signature predicts direction of nuclear repositioning

We focussed on the repositioning of the chr11 locus and investigated where this locus moved to. To identify genomic regions that were specifically contacted in response to repressor binding to the lacO, we overlapped normalized 4C profiles of each fusion protein with that of EGFP-lacR. This revealed few interactions that were induced by Nanog binding to the lacO, although one peak stood out (**Fig. 3A**). Notably, this peak coincided with a cluster of high-density Nanog binding sites [43] around the *Msi2* gene (chr11:88,152,884-88,579,581), a gene implicated in ESC self-renewal [55]. This is consistent with our previous finding that Nanog binding to the chr8 lacO array leads to subtle changes in chromosome-wide contacts specifically involving clusters of Nanog-binding sites [37]. EGFP-lacR-Ezh2 profiles also globally matched that of EGFP-lacR, except for two very prominent novel contacts with the *HoxB* cluster (chr11:96,055,675- 96,229,567) and a cluster of *Cbx* genes (chr11:118,884,343-118,947,551) (**Fig. 3B**). These loci are highly enriched for H3K27me3 (**Fig. 3B**), a modification for which Ezh2 is responsible as the enzymatic component of the Polycomb PRC2 complex [50], suggesting that Ezh2 promotes spatial crowding of genomic regions with a strong Polycomb chromatin signature.

EGFP-lacR-Suv39h1 binding to the lacO had a dramatic effect on the 4C profile, with several megabase-sized domains showing increased interaction frequencies (**Fig. 3C**). Most of these domains coincided with regions of high H3K9me3 density, the histone mark

deposited by Suv39h1 [51-53]. The chromatin flavour induced by the chromatin-protein thus determines where the target locus is repositioned to. Altogether, these data suggest that trans-acting factors promote the spatial assembly of genomic regions carrying the corresponding chromatin signature, but that genomic context determines the extent to which binding induces nuclear repositioning.

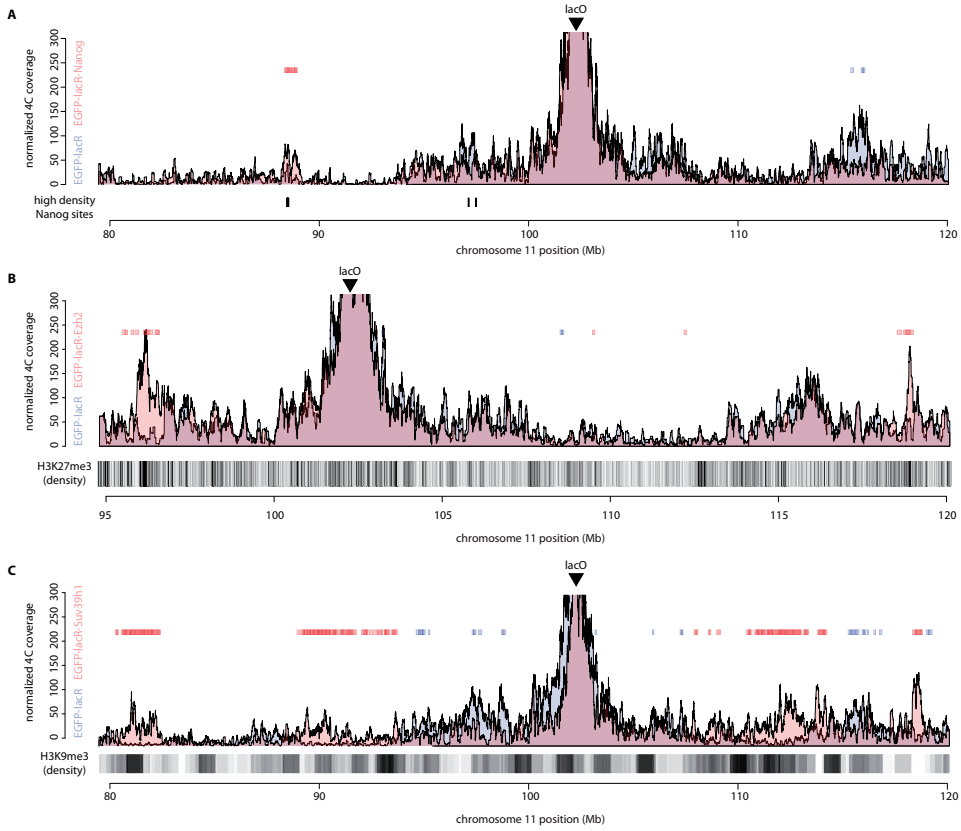


Figure 3. Local chromatin signature influences direction of nuclear repositioning

A. Comparative 4C profiles of the lacO array viewpoint from EGFP-lacR (blue) and EGFP-lacR-Nanog (red). Location of high density Nanog binding sites are shown below profiles. Red and blue bars above profiles indicate windows where 4C signal was significantly enriched or depleted, respectively, in EGFP-lacR-Nanog cells over EGFP-lacR only. Arrowheads point to the viewpoint in the lacO cassette.

B. Same as A), but for EGFP-lacR (blue) and EGFP-lacR-Ezh2 (red). Density of H3K27me3 sites is indicated below profiles.

C. Same as A), but for EGFP-lacR (blue) and EGFP-lacR-Suv39h1 (red). Density of H3K9me3 sites is indicated below profiles.

Spatial repositioning is regulated at (sub) TAD level

Having established that the chr11 lacO locus engages in contacts with other domains along its chromosome upon binding of Ezh2 and Suv39h1, we next sought to determine how focal these new contacts were around the LacO cassette. We therefore performed 4C-seq experiments with viewpoints located in newly identified Ezh2- or Suv39h1-specific contact domains. Comparative 4C profiles from the *HoxB* locus clearly discerned a peak at the position of the lacO integration site upon EGFP-lacR-Ezh2 binding, but not EGFP-lacR alone (**Fig. 4A**). Similarly, looking from a locus containing the *Asic2* gene, one of the regions contacted specifically upon Suv39h1 recruitment to LacO, comparative 4C profiles revealed a Suv39h1-specific peak at the position of the lacO transgene (**Fig. 4B**). This confirms that recruitment of these chromatin factors to the lacO locus induces new Ezh2- and Suv39h1-specific contacts. We noted that contact frequencies between the *HoxB* and *Cbx* cluster around position 119 Mb were also increased upon Ezh2 recruitment to the lacO locus (**Fig. 4A**). A similar observation was made for EGFP-lacR-Suv39h1, where genomic regions newly contacted upon Suv39h1 recruitment (**Fig. 2B**) were also more frequently contacted by the *Asic2* gene. This suggests that interactions between two loci might influence contact frequencies with other regions.

Looking more closely at the lacO locus, we noticed that the differential peak spanned a similar-sized region around the lacO integration site for both EGFP-lacR-Ezh2 and EGFP-lacR-Suv39h1 (**Fig. 4C, D**). The size of this region was reminiscent of the submegabase-scaled contact domains that were recently identified by high resolution Hi-C, which have a median size of 185 kb [14]. Indeed, the borders of the region showing new contacts corresponded to the borders of a structural domain appreciable from a Hi-C contact matrix (**Fig. 4E**) that we produced for mouse ESC cells (Geeven, Zhu et al., in prep). It therefore appears that long-range contacts induced by specific binding sites for trans-acting factors cause the entire structural domains (sub-TADs) to get engaged in new contacts with each other. From this, we conclude that spatial positioning of genomic regions is regulated at the subTAD level.

Spatial repositioning and spreading of chromatin marks are regulated on different scales

We next wished to investigate whether the topological changes were accompanied by changes in chromatin composition. Consistent with the observed spatial assembly of similarly marked chromatin regions, we found that Ezh2 and Suv39h1 recruitment led to high levels of H3K27me3 and H3K9me3 at the transgene, respectively (**Fig. 5A-C**). In case of Ezh2, this did not spread far into adjacent regions, consistent with recent findings that Ezh2 spreading upon induced recruitment of Polycomb group proteins to a tetO transgene was usually limited to <10 kb [56]. The repositioning of the lacO subTAD upon Ezh2 recruitment therefore occurs in the absence of domain-wide H3K27me3.

In contrast to Ezh2, recruitment of Suv39h1 to the lacO array did induce spreading of H3K9me3 beyond the array itself (**Fig. 5B, C**), consistent with spreading ability of classic heterochromatin [57]. The spreading was not equal on both sides, suggesting that some activity between -90kb and -60 kb one side, and +10 kb and +60 kb of the lacO array prevents spreading. Thus, heterochromatin does not spread across the whole subTAD (approx. -130

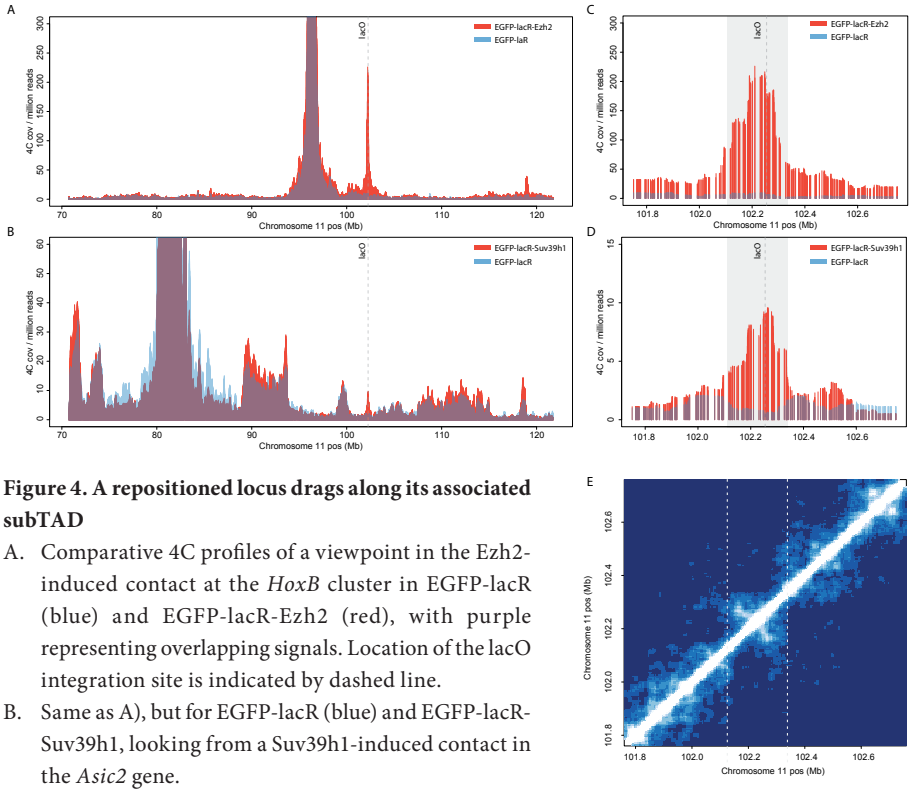


Figure 4. A repositioned locus drags along its associated subTAD

A. Comparative 4C profiles of a viewpoint in the Ezh2-induced contact at the *HoxB* cluster in EGFP-lacR (blue) and EGFP-lacR-Ezh2 (red), with purple representing overlapping signals. Location of the lacO integration site is indicated by dashed line.

B. Same as A), but for EGFP-lacR (blue) and EGFP-lacR-Suv39h1, looking from a Suv39h1-induced contact in the *Asic2* gene.

C. Same as A), zoomed in on the lacO locus. The grey area highlights the region with the highest differential 4C signal.

D. Same as B), zoomed in on the lacO locus. The grey area highlights the region with the highest differential 4C signal.

E. Hi-C contact matrix of region surrounding the lacO locus on chr11 (data from ES cells). Dashed lines correspond to the borders of the grey area in C) and D). Note that the borders coincide with edges of a block of locally increased contact frequencies ('subTAD').

till +90 kb of the lacO array) that becomes repositioned, implying that spatial repositioning and chromatin composition are regulated on different scales.

Spatial repositioning does not drive gene expression changes

Heterochromatin is usually associated with the repressive nuclear compartment. Consistently, Suv39h1 binding to the lacO led to increased contacts with regions overlapping with the gene-poor, transcriptionally inactive B compartment (**Fig. 6A**). The dramatic effect implied that a large proportion of cells moved the locus from the active A to the inactive B compartment, allowing us to look for gene expression changes in response to a nuclear compartment switch. Analysing the expression of genes in the repositioning region of chr11 by quantitative RT-PCR, we found that only the immediately flanking Rundc3a gene

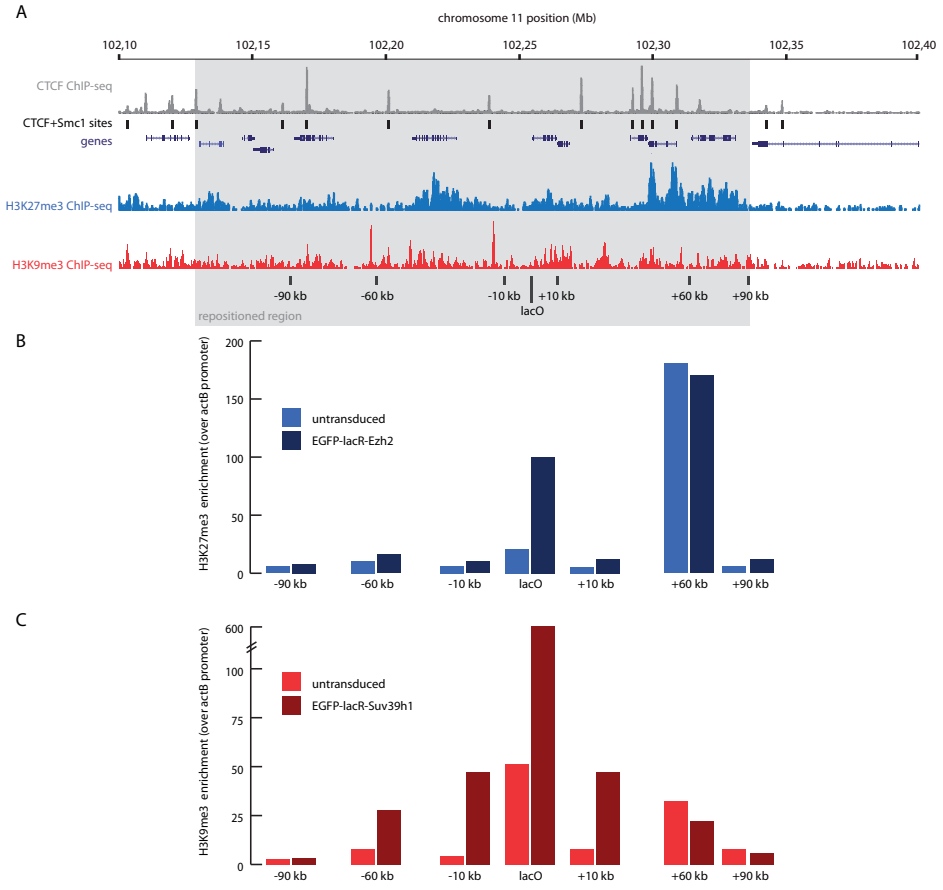


Figure 5. Induced changes in chromatin composition around lacO locus

A. Schematic overview of lacO locus on chr11, with respect to CTCF binding sites (grey), shared CTCF-cohesin sites (blocks), H3K27me3 (blue) and H3K9me3 (red) histone modifications. Locations of ChIP primers are indicated. Grey area corresponds to the subTAD that is repositioned upon Ezh2 or Suv39h1 binding (Fig. 4).

B. ChIP for H3K27me3 in untransduced (light blue) and EGFP-lacR-Ezh2 transduced (dark blue) cells for the genomic sites indicated in A). Input-normalized data is expressed as enrichment over the (H3K27me3-poor) actB promoter. Note that only the lacO array shows increased H3K27me3 levels.

C. As in B) but for H3K9me3 in untransduced (light red) and EGFP-lacR-Ezh2 (dark red) transduced cells. Note that H3K9me3 enrichment is spread outside the lacO array.

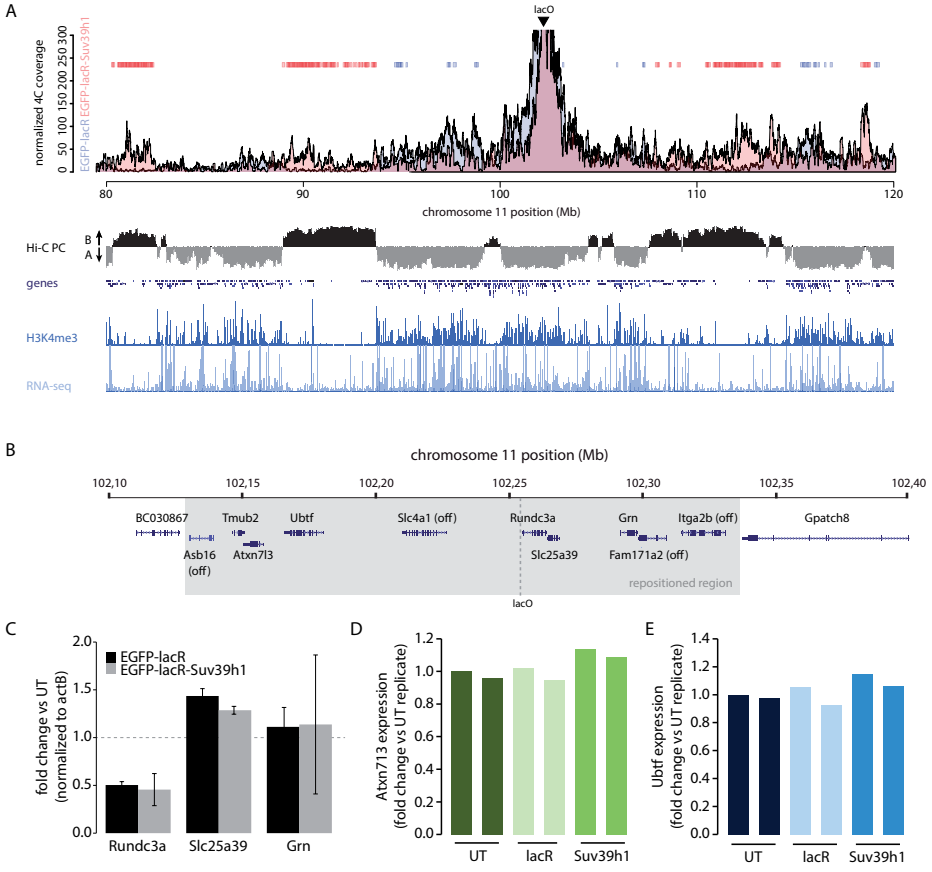


Figure 6. Spatial repositioning does not drive gene expression changes

A. Comparative 4C profiles of a viewpoint in the Suv39h1-induced contact at the *Asic2* gene in EGFP-lacR (blue) and EGFP-lacR-Suv39h1 (red), with purple representing overlapping signals. Location of the lacO integration site is indicated by dashed line. Hi-C principle component values (positive and negative corresponds to B and A compartment, respectively), genes, H3K4me3 and RNA-seq are depicted below. Red and blue bars above profiles indicate windows where 4C signal was significantly enriched or depleted, respectively, in EGFP-lacR-Suv39h1 cells over EGFP-lacR only. Note that Suv39h1-specific contacts correspond to the gene-poor and transcriptionally less active B compartment.

B. Schematic overview of genes surrounding the lacO integration site. Genes that are transcriptionally inactive in ES cells are indicated by 'off'. The grey area corresponds to the subTAD that is repositioned upon Suv39h1 binding (see also Fig. 4)

C. Quantitative RT-PCR on the 3 genes most nearby the lacO integration site. Note that the downregulated *Rundc3a* gene is only 36 bp downstream of the lacO transgene.

D. SNP-based expression analysis of the more distal *Atxn713* gene. Data plotted are the fold-changes in the fraction of reads coming from the lacO-targeted allele. Values of all replicates were normalized to the first untransduced (UT) replicate.

E. Same as D) but for *Ubtf*.

showed decreased expression levels, not only upon Suv39h1 binding but also upon LacR-GFP binding (**Fig. 6B,C**). Since the latter does not induce spatial repositioning of the domain, this repression cannot be attributed to nuclear repositioning. Genes further away and beyond the new H3K9me3 domain showed no consistent altered expression levels, not when measured by quantitative RT-PCR (**Fig. 6C** and data not shown) and not when we directly compared expression of several genes from the lacO and wild type allele by next generation sequencing of PCR amplified cDNA products harbouring allele-discriminating SNPs (**Fig. 6D-E**). We therefore conclude that spatial repositioning does not drive changes in gene expression.

DISCUSSION

The link between nuclear localisation, chromatin and gene expression is primarily based on genome-wide localization studies. Although comparisons between tissues or between differentiating cell-types have been instrumental in the analysis of the dynamics of these associations, the instructive signals mediating the observed changes are never certain. To establish cause and consequence in this three-way relationship, we therefore recruited a selection of trans-acting proteins to known genomic loci with the aim of modulating these parameters in the same cell type. Although endogenous genes are probably not targeted by so many copies of a given protein, repressive H3K27me3 and H3K9me3 marks often span genomic regions of tens of kb [58-60], not unlike the ~10 kb lacO array integrated here.

Susceptibility to spatial repositioning depends on genomic location

All three trans-acting factors were able to change the spatial positioning of the lacO locus when integrated upstream of *Rundc3a* on chr11, albeit to different extents. Whereas upon Nanog and Ezh2 binding most contacting partners elsewhere on the chromosome remained the same and only few new loci were contacted, Suv39h1 induced a dramatic repositioning of the locus to a different nuclear compartment. However, the same lacO array integrated in the *Nfix* gene on chr8 seemed impervious to these large-scale repositioning influences, showing only minor small-scale changes (**Fig. 2A**, [37]). As data from both alleles comes from the same cell with identical amounts of lacO repeats in each transgene, these differential responses cannot be explained by technical differences. This shows therefore that genomic context is an important determinant of the flexibility of a locus to roam the nucleus. This is consistent with the recently proposed dog-on-a-lead model that implies that the overall folding of a chromosome will impose restraints on the mobility and capacity of a locus to look for preferred interaction partners [36, 61]. What elements in the genomic habitat surrounding these lacO arrays are responsible for the differential capacity to re-localise is still unclear. Differences in gene density, transcriptional activity and distance to telomere and centromere have been proposed to help explain differential repositioning of the mouse and human α-globin gene upon activation in human erythoblast [62, 63], but these parameters are not markedly different between the lacO loci on chr8 and chr11 (data not shown). Studies using more systematic approaches such as TRIP [64] could perhaps help to dissect these parameters in the future.

Trans-acting factors determine nuclear position

What determines where a locus is repositioned to? Although binding of all three factors had the same repressive effect on gene expression, they guided their target locus in different directions. The observed interactions are therefore not the mere consequence of inactive loci coming together in the nucleus. Instead, it suggests that loci can have a remarkable flexibility and that silent loci can occupy different repressive subcompartments in the nucleus.

All three trans-acting factors showed a strong preference to contact genomic loci carrying a similar chromatin signature as the one deposited at the lacO array. This is consistent with the 3D clustering of Nanog-bound loci in the ESC [37], and the dominant interaction among Polycomb-bound domains [40, 41, 65]. The role of Suv39h1 in 3D chromosome conformation has not yet been investigated previously, but H3K9me3-dense chromatin was recently found to form a subcompartment in the nucleus [14]. Thus, trans-acting factors play an important role in genome organization by promoting the formation of nuclear subcompartments of specific chromatin types.

How these proteins impose the observed repositioning is not entirely clear. Given the limited freedom of chromatin to roam the interphase nucleus, repositioning probably required one or multiple rounds of mitosis [30]. Following the self-organizing principles of chromatin, chromosomes will organize into the most energetically favourable conformation [61, 66, 67]. Accordingly, after mitosis, genomic loci will have a strong preference to associate with regions with a similar chromatin type, leading to the nuclear subcompartments observed in genome-wide 3C data [6, 7, 14]. Polycomb group proteins have the intrinsic property to aggregate to establish a compact chromatin structure when bound to nearby nucleosomes [68]. Likewise, the deposition of H3K9me3 by Suv39h1 serves as a docking platform for a whole range of heterochromatic proteins, including heterochromatin protein 1 which is thought to oligomerise to promote chromatin compaction [57]. The inclination of these protein complexes to self-associate may therefore increase the probability of the lacO locus to interact with genomic regions with a similar chromatin composition in early interphase. However, we cannot exclude the possibility that new positions are adopted due to physical tethering through the simultaneous binding of the lacR and e.g. Suv39h1 moieties to the lacO and (pericentromeric) heterochromatin, respectively.

Changes in nuclear position and chromatin marks are controlled by topological constraints

As each locus is physically connected to the rest of its chromosome, its repositioning will by default carry along adjacent chromosomal regions. Differentiation-induced changes in LAD status and A/B compartments usually correspond to one or multiple neighbouring TADs [11, 19]. However, the causes and consequences of these movements are unclear; do these represent the coordinated movement of these complete TADs, or are they mostly the result of passive movement following the regulated repositioning of a single gene?

Here, we show that localised chromatin modification of a relatively small chromosomal segment induces new contacts with chromosomal regions elsewhere on the same chromosome. Reciprocally these regions now appear to not only form new contacts with the LacO array but also with the remainder of its associated subTAD. In case of Suv39h1 recruitment, the spreading

of H3K9me3 could explain the movement of part of this subTAD. However, Ezh2 binding only triggered a localized H3K27me3 domain, suggesting that the extent of spreading does not explain why the entire subTAD is pulled along with the lacO array. Thus, the nuclear position of a locus is controlled at the level of subTADs but they may adopt new positions in the nucleus because of more localized changes in their chromatin composition or associated factors.

Genomic parts of the same TAD often carry similar chromatin marks, and TAD boundaries often align well with those of chromatin domains [8, 9, 14, 36, 69-71]. Unlike classic transcription factors, writers of chromatin modifications like Ezh2 and Suv39h1 generally do not bind DNA in a sequence-specific manner, and it is unclear what governs their locus specificity given their propensity to spread particular histone marks across neighbouring nucleosomes. It was recently hypothesized that the folding of chromosomes in TADs may spatially restrict the access of locally recruited chromatin-modifying enzymes to regions in the same TAD [9].

We found that focussed recruitment of Suv39h1 to the lacO array nucleated a H3K9me3 domain that spreads over several tens of kb. However, unlike the spatial repositioning, H3K9me3 spreading did not spread across the whole subTAD. It is currently unclear what activity restricts the spreading in this region, but we note that the new H3K9me3 domain was flanked by prominent CTCF-cohesin sites (**Fig. 5**), which are conserved across all ENCODE tissues (data not shown). These sites may mark the anchors of a highly abundant loop between DNA elements that limits the reach of the histone methyltransferase. Spatial repositioning and spreading of chromatin modifications may therefore both be regulated by topological constraints, but on different scales.

Nuclear organisation is not a major determinant of gene expression

Does gene expression affect nuclear organisation? Neo expression was decreased by all recruited proteins, including the EGFP-lacR only fusion that did not induce repositioning. Thus, gene repression does not cause locus repositioning. This fits well with a recent study where they activated particular genes using TALE (transcription activator–like effector) DNA binding domains fused to a transcriptional activation domain to study the requirements for gene repositioning. They found that although transcriptional activation of endogenous genes promoted locus repositioning from the periphery to the nuclear interior, local chromatin decondensation through recruitment of an acidic peptide was also sufficient to induce similar movement [72]. This suggests that rather than transcription, the accompanying chromatin remodelling is responsible for the repositioning. This may explain why EGFP-lacR binding does not change nuclear position. It is highly conceivable that the repression of neo by EGFP-lacR recruitment is not accompanied by chromatin compaction that is known to result from Polycomb repression [50] and Suv39h1-mediated heterochromatin [57].

How about the other way around; does nuclear organisation affect gene expression? FISH studies on several developmentally regulated loci have shown that alterations in nuclear position relative to the periphery, interior or pericentromeric heterochromatin often correlate with gene expression changes [73]. Recent Hi-C data across several cell types also found a significant reduction in expression of genes that switch from the 'active' A to the 'repressive' B compartment [11]. The effect was rather small, however, and was probably caused by large

effects on only a small proportion of genes. Physical tethering of lacO arrays to the nuclear periphery affected the expression in only some cases [27, 28, 74]. It therefore remains unclear whether the repositioning of genes is a cause or consequence of gene silencing.

We found that the expression of only *neo* gene and the *Rundc3a* gene 36 bp downstream of the lacO array are affected upon repositioning of the lacO locus and the corresponding subTAD. As both genes were also repressed by EGFP-lacR binding, this cannot be attributed to changes in nuclear position. However, none of the examined neighbouring genes within the same TAD were affected by the dramatic shift from the 'active' A to the 'repressive' B compartment. Although 4C data identifies probabilistic interactions without providing information about the absolute frequency of a given interaction, the compartment switch in response to Suv39h1 binding seems quite comprehensive (**Fig. 6A**). We cannot rule out subtle changes in expression in a minority of cells, but overall our data suggests that nuclear organisation is not a major determinant of gene expression status.

This does not imply that nuclear organisation has no effect on gene expression. Rather than important for establishing gene expression status, nuclear subcompartments may aid in maintenance of expression states by creating an environment that is more or less favourable to transcriptional activation or repression in response to developmental or environmental stimuli. Based on these findings, we conclude that gene expression and higher-order chromosome topology beyond TADs are not causally related. Instead, they are independently controlled by locally associated trans-acting factors.

MATERIALS AND METHODS

Cell culture

Mouse embryonic stem cells C57Bl/6-129 were cultured as previously described [37]. In brief, they were cultured on gelatin-coated plates in BRL-conditioned DMEM (high glucose, Gibco) with 15% FBS, 1x non-essential amino acids (NEAA; Gibco), 1x penicillin–streptomycin (Gibco), 1:1000 b-mercaptoethanol (Invitrogen), 1x L-glutamine (Gibco) and 1000 U ml^{-1} leukaemia inhibitory factor (LIF; Gibco).

Generation of lacO targeted cell line.

First, the lacO array [42] was inserted in chromosome 8 (chr8)(AatII site, chr8:87321244), as described [37]. Lentiviral Cre [75] was used to delete neomycine at Chr8. Subcolonies were picked, and deletion was confirmed by PCR on genomic DNA from clonal cell lines. Then, the same lacO array was introduced into chromosome 11 (chr11:102209489) using the same gene targeting protocol. Homology arms were excised with KpnI from bacterial artificial chromosome (BAC) RP23-311P1, and the lacO array inserted at the unique AatII site. The linearized targeting construct was introduced by electroporation. After 14 days of neomycine selection, positive colonies were screened by Southern blot (Fig. 1). Probes for Southern blot were PCR-amplified from genomic DNA using the following primers:

chr8_forward: TGTGTGGTGATCATGTGTGC

chr8_reverse: TGCCACTCCTGTGTCTCAAG

chr11_forward: CCTCCTTTGGATACCTTCC
chr11_reverse: CTTTAAATCGGTGGCTGAGG

LacR-fusion constructs and Transduction

Proteins of interest were cloned in frame downstream of the EGFP-LacR fusion constructs (a gift from Pernette Verschure). To transduce EGFP-LacR fusions under control of the EF1α promoter, they were then introduced in place of the DsRed gene of the phage2-EF1α-DsRed-IRES vector (gift from Niels Geijsen) by blunt cloning. LacO cells were transduced with EGFP-lacR fusions using lentivirus based on the pHAGE2-IRES-puro backbone with an EF1α promoter [76], and selected with Puromycin (P8833, SIGMA at 1μg/μl) for approximately 10 days when cells had reached sufficient numbers for collection and tested for purity of by flow cytometry (minimum 70% GFP-positive).

Protein analysis

Nuclear extracts were made as described in [77], except for EGFP-lacR-Suv39h1, where nuclei were boiled for 5 minutes in Laemmli buffer. Immunoblot analysis were processed by using the same protocol from [37] with antibodies against GFP (ab290, Abcam) and Nanog (A300-397A, Bethyl Laboratories).

GFP distribution analysis

Stably transduced lacO cell lines were grown overnight on gelatin-coated coverslips. They were crosslinked with 4% paraformaldehyde (rT) (10 minutes), washed once with 0.125M glycine in PBS and PBS+0.1% Tween 20, and permeabilized with 0.2% Triton X-100 in PBS (5 minutes). After a final wash with PBS+0.1% Tween 20, VectaShield containing DAPI (Vector Labs) was applied and coverslips were sealed with nail polish. Images were taken with a Leica SPE confocal microscope and analyzed using ImageJ software. Maximum projections were made to allow simultaneous visualization of both lacO alleles.

Chromatin immunoprecipitation

Chromatin immunoprecipitaion (ChIP) using GFP antibody (ab290, Abcam) was done with Millipore providing protocol with a few modification. In brief, 5-10 million cells were fixed with 1% formaldehyde, then lysed in cell lysis buffer (10 mM Tris, pH 8.0; 10 mM NaCl; 0.2% NP-40; 10 mM Na-Butyrate; Proteinase inhibitor 1x) and nuclei lysis buffer (50 mM Tris, pH 8.0; 10 mM EDTA; 1% SDS; 10 mM Na-Butyrate; Proteinase inhibitor 1x). Isolated chromatin was sonicated to 500-1000 bp, processed on Covaris S2 as 30 seconds on and 30 seconds off, and then precipitated with GFP antibody overnight. After washing, chromatin was eluted, purified, and used for quantification analysis by qPCR, and data normalized to input and enrichment plotted over the untransduced cells. Histone modification ChIP was performed based on the protocol in [78] with slight modifications. In brief, 5 million cells were used in each IP. Cross-linked cells were first lysed in (50 mM Tris, pH 7.5; 150 mM NaCl; 5mM EDTA; 0.5% NP-40; 1% Triton X-100), nd sonicated to 500-1000bp in a Bioruptor (15 seconds on/15 seconds off). Antibodies used are H3K27me3 (ab6002, Abcam), H3K9me3 (ab8898, Abcam). After immunoprecipitation, isolated DNA samples were used for quantification analysis by

qPCR, and data normalized to input and enrichment plotted over the (transcriptionally active) *actB* promoter. Primers used:

genomic site	forward	reverse
actB promoter	GCAGGCCTAGTAACCGAGACA	AGTTTTGGCGATGGGTGCT
lacO	TTCGATACCTTTATCCGCTCA	GCGGATAACAATTGCTGAAG
kan^R	TGATAATCCTGATATGAATAAATTGC	TTGGCACCTTTGCTAGATTAGAA
-90 kb	ATTTTGCCTGCTGTGTGCAG	TCCCTTCTCCACAGGGACAT
-60 kb	ACTATGTGGTCTTGGCTGGC	GGAGTGGCAGGAGAGAGGTA
-10 kb	GAGTAAGCCTGACGCCTGTT	AGACACACACTGTCCTGGTG
laco	TTCGATACCTTTATCCGCTCA	GCGGATAACAATTGCTGAAG
10 kb	GACTTGCTCTCACCCCAACA	ATCTGGGTGAAAGGTGCCTG
60 kb	GCCAGCACTGACTACACCTT	GCTGGGCTGAGCAGTGAATA
90 kb	GGCTGATGGGGTGTGTTACC	ACCCACTTTCAGACGCAAGA

Gene expression analysis

RNA was isolated using TRIzol (Life technology) from cells trypsined and then converted into cDNA using random primers (Promega) using standard manufacturer providing protocol (Promega). Quantitative PCR was performed and data normalized to the actB gene. Allelic specific gene expression using high throughput sequencing was applied on cDNA synthesized in the same way. Primers used to detect gene transcripts flanked a SNP to discriminate C57Bl/6 allele from the 129S1/SvImJ allele. PCR product amplified from cDNA was sequenced on Illumina GAII. Sequencing reads were mapped to each allele to determine the fraction of reads coming from the lacO-transgenic Bl/6 allele. Fold-changes were expressed relative to the Bl/6 fraction in untransduced cells.

Primer used:

genomic site	forward	reverse
Hprt	TCCTCCTCAGACCGCTTTT	CCTGGTTCATCATCGCTAATC
neo set-1	ATGCCTGCTTGCCGAATA	CCACAGTCGATGAATCCAGA
neo set-2	GCAGGATCTCCTGTCATCTCA	TAGCCGGATCAAGCGTATG
Grn	GCCCGTTCTCTAAGGGTGTG	ACAGCACCCAAGGGGTTATC
Slc25a39	AGGCAGTATCTTGGCCCCAT	GCACACGTACCCCAAGACA
mRundc3a	AAGGGCGAAGTTCTGGATGG	GGTAGTCGTAGCTTTGGGTGA
Atxn7l3	GGTTTGGGGCTCTGAGGAAA	TGGGAGGTGGGATACAGGTC
Ubtf	GCTCCTCTAACTGCTTGCCA	GGAGAGCCTACTTCCCACCT

4C sequence analysis

Standard 4C experiments were done as previously described [47]. For allele specific 4C we have used a paired-end 4C strategy [79], where the forward primer analyses the ligation product and the reverse primer flanks a SNP. After sequencing, this SNP is used to demultiplex the two alleles, to create two separate 4C profiles. Allelic specific 4C from the chr8 lacO viewpoint and reciprocal single-end 4C used a combination of HindIII and DpnII, single-end 4C from chr11 lacO viewpoint used a combination of HindIII and NlaIII. 4C data mapping and analysis was as described [37]. Domainograms were produced as described before [37]. To allow direct comparisons of 4C profiles of different samples, mapped reads from each sample were normalized for sequencing depth by scaling with a constant to force equal numbers of mapped reads on the entire bait- or cis-chromosome. Subsequent 4C-Seq coverage plots thus show the 4C coverage as rolling mean averages of mapped reads in windows of a fixed number of 4C fragments, per one million of mapped reads. To compare 4C-Seq coverage between 2 conditions in a genomic region defined by any number of consecutive 4C fragments, we first binarized the 4C mapped reads at 4C fragments to distinguish between fragments that were captured and those that were not captured. Next, for each given region, we summed the total number of captured and not captured fragments in that region in both conditions to create a 2x2 table of counts. Then we tested the null hypothesis that the probability of capturing fragments is not different in the given region with Fisher's exact test and reject if the corresponding p-value exceeds. Primers used:

allelic Paired-end	**primer 1**	**primer 2**
chr8	AATGATACGGCGACCACCGAGATCTA-CACTCTTTCCCTACACGACGCTCTTC-CGATCTCTGGAACTAAATGGAGGATC	CAAGCAGAAGACGGCATACGAGA-TCGGTCTCGGCATTCCTGCTGAAC-CGCTCTTCCGATCTCAAGCAGAA-GACGGCATACGAGATCGGTCTCG-GCATTCCTGCTGAACCGCTCTTC-CGATCTTACCAGGACCCCTGGGACCC
single-end (viewpoint)	**primer 1**	**primer 2**
neo	AATGATACGGCGACCAC-CGAACACTCTTTCCCTACAC-GACGCTCTTCCGATCTCGAAGTTATC-GATCGAAGCTT	CAAGCAGAAGACGGCATACGAA-GAAAAGCGGCCATTTTCCA
HoxB	AATGATACGGCGACCAC-CGAACACTCTTTCCCTACAC-GACGCTCTTCCGATCTTCCCCTGGAT-GAGGAAGCTT	CAAGCAGAAGACGGCATACGAGA-GCGGTTGACGCTGAGATC
Asci2	AATGATACGGCGACCAC-CGAACACTCTTTCCCTACAC-GACGCTCTTCCGATCTTGTTAGGTG-GCACCAAGCTT	CAAGCAGAAGACGGCATACGAAC-TAATGATGGGCACAGTTT

For comparative analysis, Hi-C and principle component analysis data from embryonic stem cells was taken from Geeven, Zhu et al., (in prep), density maps of Nanog from [43], Smc1 ChIP-seq data from [80], RNA-seq and CTCF, H3K4me3, H3K27me3 and H3K9me3 ChIP seq data from the ENCODE project [81] .

Acknowledgements: We thank Daan Noordermeer for making targeting constructs. Niels Geijssen for sharing phage2 construct. Erik Splinter, Petra Klous and Harmen van de werken for technical support. Stefan van der Elst for FACS sorting, Stieneke van den Brink for ES cell culture, Anko de Graaff for imaging facility.

REFERENCES

1. Lanzuolo, C., et al., *Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex.* Nat Cell Biol, 2007. **9**(10): p. 1167-74.
2. Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. **295**(5558): p. 1306-11.
3. de Wit, E. and W. de Laat, *A decade of 3C technologies: insights into nuclear organization.* Genes Dev, 2012. **26**(1): p. 11-24.
4. Bickmore, W.A. and B. van Steensel, *Genome architecture: domain organization of interphase chromosomes.* Cell, 2013. **152**(6): p. 1270-84.
5. Bolzer, A., et al., *Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes.* PLoS Biol, 2005. **3**(5): p. e157.
6. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* Nat Genet, 2006. **38**(11): p. 1348-54.
7. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.
8. Dixon, J.R., et al., *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature, 2012. **485**(7398): p. 376-80.
9. Nora, E.P., et al., *Spatial partitioning of the regulatory landscape of the X-inactivation centre.* Nature, 2012. **485**(7398): p. 381-5.
10. Shen, Y., et al., *[TALE nuclease engineering and targeted genome modification].* Yi Chuan, 2013. **35**(4): p. 395-409.
11. Dixon, J.R., et al., *Chromatin architecture reorganization during stem cell differentiation.* Nature, 2015. **518**(7539): p. 331-6.
12. Tanay, A. and G. Cavalli, *Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization.* Curr Opin Genet Dev, 2013. **23**(2): p. 197-203.
13. Zhang, Y., et al., *Spatial organization of the mouse genome and its role in recurrent chromosomal translocations.* Cell, 2012. **148**(5): p. 908-21.
14. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell, 2014. **159**(7): p. 1665-80.
15. Guelen, L., et al., *Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions.* Nature, 2008. **453**(7197): p. 948-51.
16. Nemeth, A., et al., *Initial genomics of the human nucleolus.* PLoS Genet, 2010. **6**(3): p. e1000889.
17. van Koningsbruggen, S., et al., *High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli.* Mol Biol Cell, 2010. **21**(21): p. 3735-48.
18. Zink, D., et al., *Transcription-dependent spatial arrangements of CFTR and adjacent genes in human cell nuclei.* J Cell Biol, 2004. **166**(6): p. 815-25.

19. Peric-Hupkes, D., et al., *Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation.* Mol Cell, 2010. **38**(4): p. 603-13.
20. Brown, K.E., et al., *Association of transcriptionally silent genes with Ikaros complexes at centromeric heterochromatin.* Cell, 1997. **91**(6): p. 845-54.
21. Ragoczy, T., et al., *The locus control region is required for association of the murine beta-globin locus with engaged transcription factories during erythroid maturation.* Genes Dev, 2006. **20**(11): p. 1447-57.
22. Williams, R.R., et al., *Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus.* J Cell Sci, 2006. **119**(Pt 1): p. 132-40.
23. Zullo, J.M., et al., *DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina.* Cell, 2012. **149**(7): p. 1474-87.
24. Chubb, J.R., et al., *Chromatin motion is constrained by association with nuclear compartments in human cells.* Curr Biol, 2002. **12**(6): p. 439-45.
25. Chuang, C.H., et al., *Long-range directional movement of an interphase chromosome site.* Curr Biol, 2006. **16**(8): p. 825-31.
26. Khanna, N., Y. Hu, and A.S. Belmont, *HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation.* Curr Biol, 2014. **24**(10): p. 1138-44.
27. Reddy, K.L., et al., *Transcriptional repression mediated by repositioning of genes to the nuclear lamina.* Nature, 2008. **452**(7184): p. 243-7.
28. Finlan, L.E., et al., *Recruitment to the nuclear periphery can alter expression of genes in human cells.* PLoS Genet, 2008. **4**(3): p. e1000039.
29. Kind, J., et al., *Single-cell dynamics of genome-nuclear lamina interactions.* Cell, 2013. **153**(1): p. 178-92.
30. Kumaran, R.I. and D.L. Spector, *A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence.* J Cell Biol, 2008. **180**(1): p. 51-65.
31. Fraser, P. and W. Bickmore, *Nuclear organization of the genome and the potential for gene regulation.* Nature, 2007. **447**(7143): p. 413-7.
32. Bhan, S., et al., *Global gene expression analysis reveals specific and redundant roles for H1 variants, H1c and H1(0), in gene expression regulation.* Gene, 2008. **414**(1-2): p. 10-8.
33. Ayyanathan, K., et al., *Regulated recruitment of HP1 to a euchromatic gene induces mitotically heritable, epigenetic gene silencing: a mammalian cell culture model of gene variegation.* Genes Dev, 2003. **17**(15): p. 1855-69.
34. Palstra, R.J., et al., *Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription.* PLoS One, 2008. **3**(2): p. e1661.
35. Hakim, O., et al., *Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements.* Genome Res, 2011. **21**(5): p. 697-706.
36. Noordermeer, D., et al., *Variegated gene expression caused by cell-specific long-range DNA interactions.* Nat Cell Biol, 2011. **13**(8): p. 944-51.
37. de Wit, E., et al., *The pluripotent genome in three dimensions is shaped around pluripotency factors.* Nature, 2013. **501**(7466): p. 227-31.
38. Schoenfelder, S., et al., *Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells.* Nat Genet, 2010. **42**(1): p. 53-61.
39. Xu, M. and P.R. Cook, *Similar active genes cluster in specialized transcription factories.* J Cell Biol, 2008. **181**(4): p. 615-23.
40. Tolhuis, B., et al., *Interactions among Polycomb domains are guided by chromosome architecture.* PLoS Genet, 2011. **7**(3): p. e1001343.
41. Bantignies, F., et al., *Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila.* Cell, 2011. **144**(2): p. 214-26.
42. Lau, I.F., et al., *Spatial and temporal organization of replicating Escherichia coli chromosomes.* Mol Microbiol, 2003. **49**(3): p. 731-43.

43. Marson, A., et al., *Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells.* Cell, 2008. **134**(3): p. 521-33.
44. Cheutin, T. and G. Cavalli, *Polycomb silencing: from linear chromatin domains to 3D chromosome folding.* Curr Opin Genet Dev, 2014. **25**: p. 30-7.
45. Aagaard, L., et al., *Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31.* EMBO J, 1999. **18**(7): p. 1923-38.
46. Melcher, M., et al., *Structure-function analysis of SUV39H1 reveals a dominant role in heterochromatin organization, chromosome segregation, and mitotic progression.* Mol Cell Biol, 2000. **20**(10): p. 3728-41.
47. Splinter, E., et al., *Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation.* Methods, 2012. **58**(3): p. 221-30.
48. Splinter, E., et al., *The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA.* Genes Dev, 2011. **25**(13): p. 1371-83.
49. Liang, H.L., et al., *The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila.* Nature, 2008. **456**(7220): p. 400-3.
50. Simon, J.A. and R.E. Kingston, *Mechanisms of polycomb gene silencing: knowns and unknowns.* Nat Rev Mol Cell Biol, 2009. **10**(10): p. 697-708.
51. Bannister, A.J., et al., *Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain.* Nature, 2001. **410**(6824): p. 120-4.
52. Lachner, M., et al., *Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins.* Nature, 2001. **410**(6824): p. 116-20.
53. Nakayama, J., et al., *Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly.* Science, 2001. **292**(5514): p. 110-3.
54. Dubarry, M., et al., *Tight protein-DNA interactions favor gene silencing.* Genes Dev, 2011. **25**(13): p. 1365-70.
55. Zhao, R., M.S. Bodnar, and D.L. Spector, *Nuclear neighborhoods and gene expression.* Curr Opin Genet Dev, 2009. **19**(2): p. 172-9.
56. Blackledge, N.P., et al., *Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation.* Cell, 2014. **157**(6): p. 1445-59.
57. Grewal, S.I. and S. Jia, *Heterochromatin revisited.* Nat Rev Genet, 2007. **8**(1): p. 35-46.
58. Bernstein, B.E., et al., *A bivalent chromatin structure marks key developmental genes in embryonic stem cells.* Cell, 2006. **125**(2): p. 315-26.
59. Pauler, F.M., et al., *H3K27me3 forms BLOCs over silent genes and intergenic regions and specifies a histone banding pattern on a mouse autosomal chromosome.* Genome Res, 2009. **19**(2): p. 221-33.
60. Hawkins, R.D., et al., *Distinct epigenomic landscapes of pluripotent and lineage-committed human cells.* Cell Stem Cell, 2010. **6**(5): p. 479-91.
61. Krijger, P.H. and W. de Laat, *Identical cells with different 3D genomes; cause and consequences?* Curr Opin Genet Dev, 2013. **23**(2): p. 191-6.
62. Brown, J.M., et al., *Coregulated human globin genes are frequently in spatial proximity when active.* J Cell Biol, 2006. **172**(2): p. 177-87.
63. Brown, J.M., et al., *Association between active genes occurs at nuclear speckles and is modulated by chromatin environment.* J Cell Biol, 2008. **182**(6): p. 1083-97.
64. Akhtar, W., et al., *Chromatin position effects assayed by thousands of reporters integrated in parallel.* Cell, 2013. **154**(4): p. 914-27.
65. Denholtz, M., et al., *Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization.* Cell Stem Cell, 2013. **13**(5): p. 602-16.
66. Misteli, T., *Beyond the sequence: cellular organization of genome function.* Cell, 2007. **128**(4): p. 787-800.

67. Gibcus, J.H. and J. Dekker, *The hierarchy of the 3D genome.* Mol Cell, 2013. **49**(5): p. 773-82.
68. Francis, N.J., R.E. Kingston, and C.L. Woodcock, *Chromatin compaction by a polycomb group protein complex.* Science, 2004. **306**(5701): p. 1574-7.
69. Sexton, T., et al., *Three-dimensional folding and functional organization principles of the Drosophila genome.* Cell, 2012. **148**(3): p. 458-72.
70. Hou, Z., et al., *Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis.* Proc Natl Acad Sci U S A, 2013. **110**(39): p. 15644-9.
71. Andrey, G., et al., *A switch between topological domains underlies HoxD genes collinearity in mouse limbs.* Science, 2013. **340**(6137): p. 1234167.
72. Therizols, P., et al., *Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells.* Science, 2014. **346**(6214): p. 1238-42.
73. Ferrai, C., et al., *Gene positioning.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a000588.
74. Kumaran, R.I., R. Thakar, and D.L. Spector, *Chromatin dynamics and gene positioning.* Cell, 2008. **132**(6): p. 929-34.
75. Splinter, E., et al., *CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus.* Genes Dev, 2006. **20**(17): p. 2349-54.
76. Wilson, A.A., et al., *Sustained expression of alpha1-antitrypsin after transplantation of manipulated hematopoietic stem cells.* Am J Respir Cell Mol Biol, 2008. **39**(2): p. 133-41.
77. Andrews, N.C. and D.V. Faller, *A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells.* Nucleic Acids Res, 1991. **19**(9): p. 2499.
78. Schmidt, D., et al., *ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions.* Methods, 2009. **48**(3): p. 240-8.
79. Holwerda, S.J., et al., *Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells.* Nucleic Acids Res, 2013. **41**(14): p. 6905-16.
80. Kagey, M.H., et al., *Mediator and cohesin connect gene expression and chromatin architecture.* Nature, 2010. **467**(7314): p. 430-5.
81. ENCODE Project Consortium, *A user's guide to the encyclopedia of DNA elements (ENCODE).* PLoS Biol, 2011. **9**(4): p. e1001046.

SUPPLEMENTARY MATERIALS

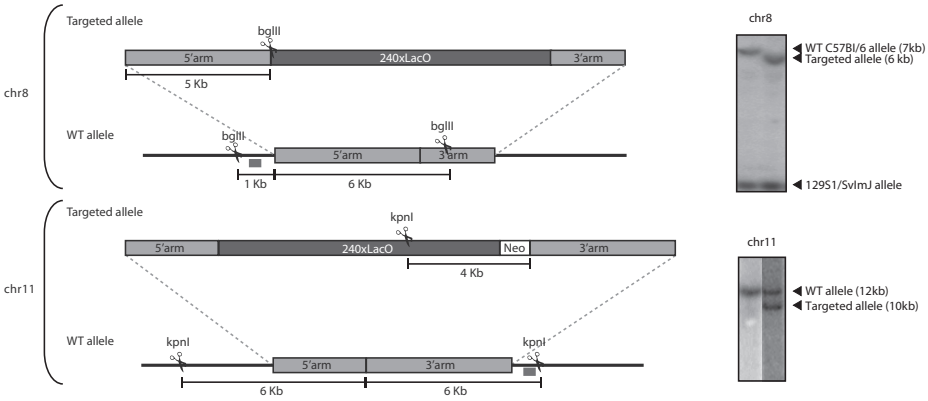


Figure S1. Validation of lacO integration sites by Southern blot
Strategy and data from Southern blot analysis confirming correct integration of the lacO cassette on chr8 (top) and chr11 (bottom). Red blocks indicate probes used for hybridization.

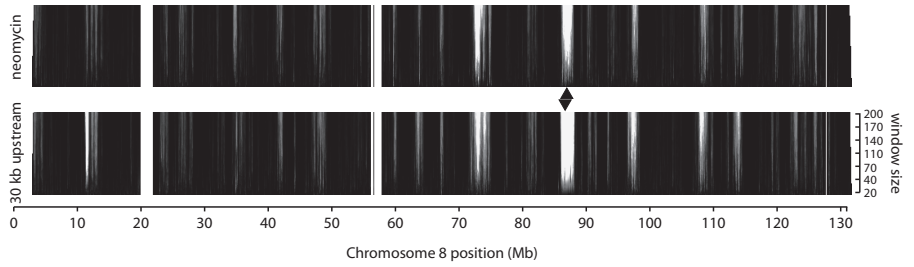


Figure S2. a viewpoint 30 kb upstream of the chr8 lacO array produces profiles similar to that of the neomycin cassette
Chromosome 8 domainograms showing allele-specific 4C profiles on the transgenic BI/6 allele comparing the 'neomycin' and '30 kb upstream' viewpoints in untransduced ES cells carrying only a lacO transgene on chr8, with the neomycin gene still intact. Viewpoints are indicated by arrowheads. Note that the profiles are highly similar.

The pluripotent genome in three dimensions is shaped around pluripotency factors

Elzo de Wit[1*], Britta A. M. Bouwman[1*], Yun Zhu[1], Petra Klous[1], Erik Splinter[1], Marjon J. A. M. Verstegen[1], Peter H. L. Krijger[1], Nicola Festuccia[2], Elphe`ge P. Nora[3], Maaike Welling[1], Edith Heard[3], Niels Geijsen[1,4], Raymond A. Poot[5], Ian Chambers[2] & Wouter de Laat[1]

[1] Hubrecht Institute-KNAW & University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.
[2] MRC Centre for Regenerative Medicine, Institute for Stem Cell Research, School of Biological Sciences, University of Edinburgh, 5 Little France Drive, Edinburgh EH16 4UU, UK.
[3] Mammalian Developmental Epigenetics Group, Institut Curie CNRS UMR3215 INSERM U934, Paris, France.
[4] Utrecht University, School of Veterinary Medicine, 3584 CL Utrecht, The Netherlands.
[5] Department of Cell Biology, Erasmus Medical Center, Rotterdam 3015 GE, The Netherlands.
*These authors contributed equally to this work.

 # Adapted from Nature 2013. 501(7466): p. 227-31.

ABSTRACT

It is becoming increasingly clear that the shape of the genome importantly influences transcription regulation. Pluripotent stem cells (PSCs) such as embryonic stem cells (ESCs) were recently shown to organize their chromosomes into topological domains that are largely invariant between cell types[1,2]. Here, we applied 4C technology and combined ChIP-seq with Hi-C data to demonstrate that inactive chromatin is unusually disorganized in PSC nuclei. We show that gene promoters engage in contacts between topological domains in a largely tissue-independent manner while enhancers have a more tissue-restricted interaction profile. Most strikingly, genomic clusters of pluripotency factor binding sites find each other very efficiently, in a manner that is strictly PSC-specific, dependent on the presence of Oct4 and Nanog protein and inducible upon artificial recruitment of Nanog to a selected chromosomal site. We conclude that pluripotent stem cells have a unique higher-order genome structure shaped by pluripotency factors. We speculate that this interactome enhances the robustness of the pluripotent state.

In recent years, several technological advances have made it possible to delineate the three-dimensional shape of the genome[3]. Spatial organization of DNA has been recognized as an additional regulatory layer of chromatin, important for gene regulation and transcriptional competence[4, 5]. In somatic cells active and inactive chromosomal regions are spatially segregated[6, 7]. Recently, the genome was further shown to be subdivided into evolutionarily conserved topological domains[1, 2].

4C is a genome-scale variant of the 3C technology[8], which examines the spatial organization of DNA and measures the contact frequencies of a chosen genomic site, or "viewpoint", with the rest of the genome. To assess chromosome topology in mouse E14 ESCs (IB10), we generated high-resolution contact maps using 4C-seq[9] for a series of individual sites representative of different chromosomal regions on various chromosomes (see Supplementary Fig. 1 and Supplementary Methods). All 4C experiments show the typical result of a chromosome conformation capture experiment, with the bulk of the signal close to the viewpoint, intrachromosomal captures outnumbering interchromosomal captures, and clustering of captures at distal sites[6, 7] (Supplementary Fig. 2 and Supplementary table 1). To identify significant intra- and interchromosomal contacts, we used a windowing approach in combination with a false discovery rate analysis that determines significant clustering of independently captured sequences (FDR, $\alpha = 0.01$ ref:[10]). Contacts in this case can mean either direct interactions between the chromatin of chromosomal regions or indirect contacts via intermediate protein complexes. 3D-DNA FISH experiments validated the 4C results (Supplementary Fig. 3).

Different from what is observed in somatic cells, in ESCs we find that transcriptionally inactive regions form low numbers of specific long-range contacts (Figure 1a, b). This is not due to their inability to reach over large distances, but rather to a more random organization of their long-range captures (Supplementary Fig. 4), suggesting that inactive chromatin is spatially less organized in pluripotent nuclei. We confirmed these results in an independent, 129/Cast, ESC line[11] (Supplementary Figure 5a). We furthermore show that this is not an intrinsic feature of the selected regions as they do engage in many long-range contacts in astrocytes (Fig. 1c and Supplementary Fig. 6). For example the chemoreceptor *Tas2R110* gene, part of a cluster of taste receptors that is specifically expressed in taste buds, engages in only 3 contacts in ESCs but shows 34 specific contacts in astrocytes (Figure 1d).

We assessed whether the lack of long-range contacts is a global feature of ESCs, by analyzing a recently published mouse ESC Hi-C dataset[1]. "Virtual 4C" contact profiles extracted from the Hi-C dataset (see Methods for details) correlate strongly to our "true 4C" profiles (Supplementary Fig. 7), emphasizing the high level of agreement between the datasets. The Hi-C data confirm on a global scale that inactive and active chromatin differ in their propensity to form specific long-range contacts in ESCs. Similar to our 4C data, this difference is abolished in differentiated cells (cortex) where both active and inactive chromatin engage equally well in specific long-range contacts (Figure 1e, f, Supplementary Fig. 8).

We next asked whether chromosomal organization is reversed during cellular reprogramming. 129/Cast NPCs[9] were transduced with a lentivirus containing a multicistronic transcript encoding Oct4, Klf4, Sox2 and c-Myc, to generate induced

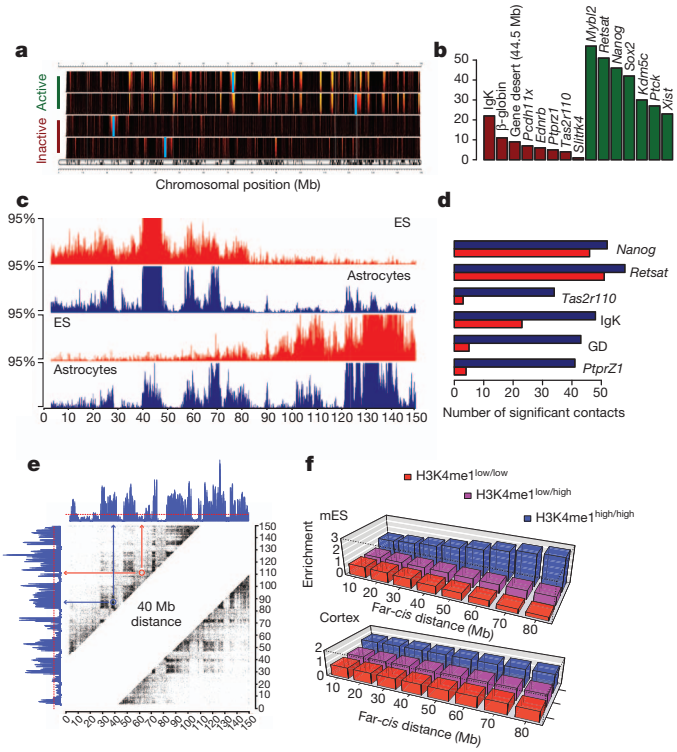


Figure 1. Inactive regions lack specific long-range interactions in embryonic stem cells A) Domainogram analysis (see Methods) shows 4C profiles of *Retsat*, *Nanog*, *Ptprz1* and gene desert in ESCs, showing the differential profiles for active (darkgreen) and inactive (darkred) viewpoints. Below the domainograms a map of the chromosomal position of the genes is plotted; top shows genes on the positive strand, bottom genes on the negative strand. B) Quantification of the number of significantly contacted regions for different viewpoints in ESCs. Green bars denote viewpoints in active regions, red bars denote viewpoints in inactive regions. C) Chromosomal maps show the distribution of read counts for a gene desert (at 44.5Mb) and for *Tas2R110,* for ESCs (red) and astrocytes (blue). 4C signal is calculated using a sliding window average (running mean) of the read counts (window size is 51). The vertical axis is maximized at the 95th percentile. D) Quantification of the number of far-*cis* regions that are significantly contacted by a given viewpoint in ESCs (red) or astrocytes (blue). E) A pairwise contact matrix was generated to calculate disorganization scores from the HiC data (see Methods). Chromosome 6 was segmented into regions with high density of H3K4me1 and low density of H3K4me1, as a proxy for active and inactive chromatin. The pairwise contact matrix was subdivided into contacts between two regions of high H3K4me1 density (H3K4me1$^{high/high}$) or low H3K4me1 density (H3K4me1$^{low/low}$) or between a region with low H3K4me1 density and a region with high H3K4me1 density (H3K4me1$^{low/high}$). F) From the distribution of H3K4me1 high and low regions we calculated an expected distribution of long-range contacts, under the null hypothesis that there is no difference between active and inactive regions with respect to their long-range contacts. An enrichment score is calculated by dividing the observed scores by these expected values.

pluripotent stem cells (iPSCs). qPCR expression analysis of several marker genes confirmed reprogramming (Supplementary Fig. 9). A reactivated gene (*Nanog*) gains contacts during iPSC reprogramming, whereas a resilenced gene (*Ptprz1*) loses all but two contacts (Supplementary Fig. 10), demonstrating that cellular reprogramming is accompanied by the reemergence of a pluripotency-specific spatial organization of the genome.

A closer inspection of the intrachromosomal contacts made by the *Nanog* gene revealed another aspect of the 3D pluripotent genome; *Nanog* was found to interact with many genes that are known to play an important role in maintenance of ESC pluripotency, including *Rybp*, *Ezh2*, *Tcf3* and *Smarcad1*. The ESC-specific nature of these contacts becomes obvious from a comparison between the ESC and NPC contact profiles of the *Nanog* gene. Most of the ESC-specific interacting regions have a high density of binding sites for the pluripotency factors Oct4/Sox2/Nanog (Fig. 2a, b). Importantly, we also find such preferential associations when we apply 4C to the *Sox2* enhancer (Supplementary Fig. 11) The ESC Hi-C data[1] show that *Nanog*-contacting regions also form preferential contacts amongst each other (Fig. 2c). Among the interchromosomal contacts made by the *Nanog* gene are again a large number of pluripotency related genes: *Mybl2*, *Dppa5*, *Rex1*, *Zfp281*, *Lefty1*, *Lin28a*, *Esrrb, Klf5, Sall1*, *Cbx5* and *Cbx7* (Figure 2d, Supplementary Fig. 12, see Supplementary Table 2 for a full list of contacted regions). The contacts of *Nanog* with *Esrrb* and *Zfp281* were verified by 3D-DNA FISH (Supplementary Fig. 12). GREAT analysis[12] of these interchromosomal contacts reveals strong enrichment for genes involved in pluripotency and early embryogenesis, which is not observed for unrelated viewpoints or in other tissues (Supplementary Table 3). This suggests that pluripotency genes prefer to cluster with other pluripotency-specific genes.

We designed a computational strategy, PE-SCAn (Fig. 2e), which combines ChIP-seq and Hi-C data to analyze, on a global scale, whether given genomic sites (e.g. sites bound by a protein of interest) in different topological domains have a preference to interact amongst each other. PE-SCAn shows that individual Nanog, Sox2, or Oct4 binding sites have little preference to contact each other over such large chromosomal distances (Supplemental Fig. 13a). However, clusters of Nanog, Oct4, or Sox2 binding sites (5 or more per 50 kb) do show a strong preference to interact with each other in ESCs (Fig. 2f). When we circularly permute the positions of the Nanog, Sox2, or Oct4 clusters, this preference is not observed, confirming that these interactions are specific (data not shown). Moreover, these contacts are tissue-specific as they are absent in cortex (Fig 2f).

We also used PE-SCAn to investigate the contribution of other factors to the shape of the pluripotent genome. While CTCF and cohesin have both been implicated in higher-order chromatin folding, CTCF has been suggested to predominantly form chromatin loops over shorter distances[13, 14]. Indeed, we find that CTCF, but also cohesin binding sites, contribute little to chromosome folding over larger distances (Fig. 3a and Supplementary figure 13b). Recent chromosome architecture experiments have revealed a central role for promoters in chromosome topology[15 Snyder]. PE-SCAn for H3K4me3 confirmed that active transcriptional start sites are engaged in specific long-range contacts (Fig. 3b). However, their contribution is largely tissue-invariant, because promoters marked by H3K4me3 in either ESC or cortex also find each other equally well in the corresponding tissue (Fig. 3b). This is different for active

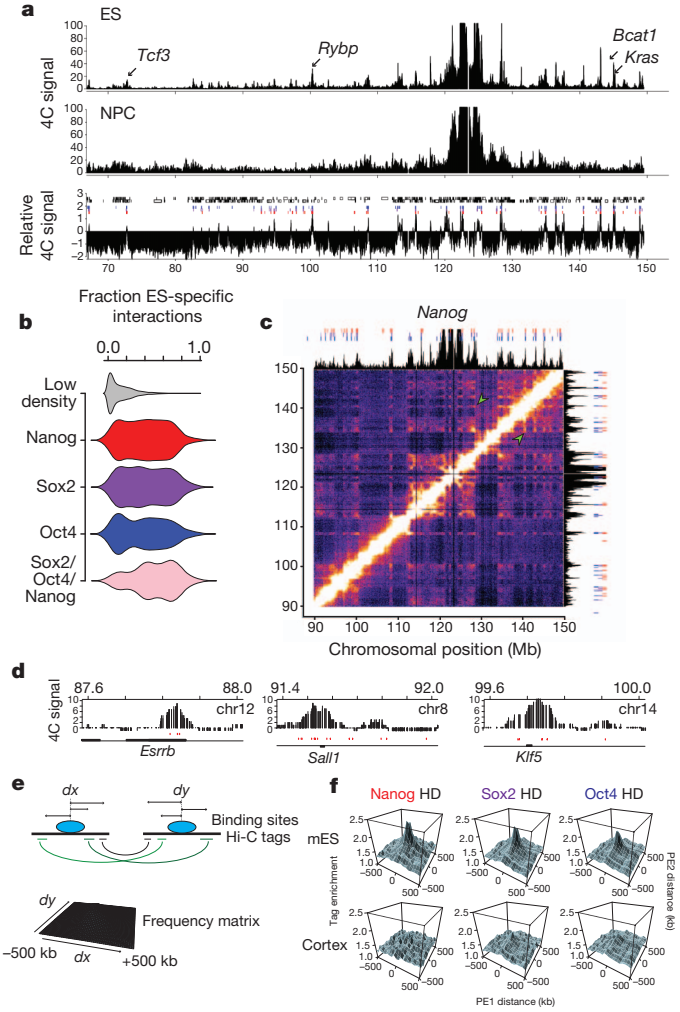


Figure 2. Expressed *Nanog* gene shows preferential interaction with other pluripotency genes. A) Chromosomal map of 4C signal for the *Nanog* gene in ESCs and NPCs. 4C data is normalized to reads per million and plotted as a running mean with a window of 31. Bottom panel shows the ESC over NPC ratio. Red, purple and blue rectangles denote the windows with a high density of Nanog, Sox2, and Oct4 binding sites, respectively. High-density is defined as > 5 sites/100 kb. B) Violin plots show quantification of ESC/NPC ratios for regions with a high density of binding sites for Nanog, Sox2, Oct4 and all three combined. C) Combined HiC-4C plot for the telomeric region of chromosome 6, shows a normalized HiC contact matrix (see Methods) with the 4C data for *Nanog* superimposed. Red, purple and blue rectangles show the high-density regions like in A) Green arrowheads point to Nanog HD Hi-C interactions other than with the *Nanog* enhancer. D) Examples of interchromosomal contacts made by *Nanog* with pluripotency genes *Esrrb*[26], *Sall1*[27] and *Klf5*[28]. See methods for the definition of 4C enrichment score. Nanog, Sox2 and Oct4 binding sites are again highlighted with red, purple and blue rectangles. E) Schematic depiction of Paired-End Spatial Chromatin Analysis (PE-SCAn). ▶

enhancer sites (H3K27ac[16]), which contribute to genome topology in a more tissue-restricted manner (Fig 3b). Pluripotency factors, but also cohesin, often bind to enhancer sequences. For Oct4, Sox2, Nanog and cohesin we find that 41%, 38%, 35% and 27% of binding sites overlap with active enhancer sites. All intersected enhancer sites show an equal preference for homo-typic contacts as the unselected enhancers (Fig. 3c). Importantly, the preferred contacts among *Nanog*-enhancers were not dependent on cohesin (Fig. 3d). Finally, we assessed chromosomal contacts among, respectively, enhancer and cohesin clusters (5 or more per 50 kb). We found that they have no advantage over isolated sites to interact with each other, and that their contact preference is not as pronounced as seen for clusters of pluripotency factor binding sites (Supplementary Fig. 13b).

To investigate whether this pluripotency-specific genome configuration is dependent on pluripotency factors, we used ZHBTc4[17] and RCNβH[18] ES cell lines, which allow the acute depletion of Oct4 protein or Nanog protein, respectively (Figure 4a,b, and Supplementary Figure 14a,b). After Oct4 or Nanog protein removal, the overall chromosome topology is largely unaffected (Figure 4c and Supplementary Figure 14c). However, a close comparison between factor-depleted and wild type contact profiles reveals a decrease in contact frequencies specifically at clusters where pluripotency factors normally bind (Fig. 4d,e). Quantification confirms that the regions with reduced contact frequency after removal of Oct4 or Nanog protein are those with a high density of cognate binding sites and not, for example, regions with a high density of CTCF binding sites (Figure 4f, see Supplementary Methods for details). Of note, partial loss of Nanog by siRNA-mediated knock-down (78%) has no effect (Supplementary Fig. 15), implying that full knock-out of Nanog protein is required to affect chromosome topology in ESCs.

To further test whether pluripotency factor binding has a direct role in this PSC-specific genome configuration we made use of a C56Bl/6-129S1/SvImJ ES cell line with a 256x lacO repeat cassette integrated into the C56Bl/6 *Nfix* allele on chromosome 8 (Figure 4g and Supplementary Fig. 16). We targeted GFP-LacR-Nanog fusion proteins to these lacO repeats and performed allele-specific 4C[19] to simultaneously analyze the contact spectra of the targeted C56Bl/6 and the non-targeted 129S1/SvImJ allele. Again, the overall chromosome topology for both alleles was highly similar, but a number of new specific contacts were found for the C56Bl/6 allele. Strikingly, these contacts coincide with high-density Nanog binding sites around the pluripotency genes *Sall1* and *Klf2* and the *Irx* cluster of developmental regulators. Circular permutation of the positions of the high-density Nanog clusters showed that increased contact frequency was significantly enriched at these sites ($p << 0.001$),

▶ HiC di-tags are sequentially aligned to ChIPseq binding sites. From the total set of distances *dx* and *dy* a normalized two-dimensional frequency matrix is calculated (see Methods). Hi-C pairs within 5 megabases were excluded to focus the analysis on contacts between, rather than within topological domains. F) PE-SCAn plots show the alignment of intrachromosomal Hi-C data to high-density clusters of pluripotency factors (5 sites in 50kb). Top row shows the alignment of ESC Hi-C data, bottom row shows alignment of cortex Hi-C data.

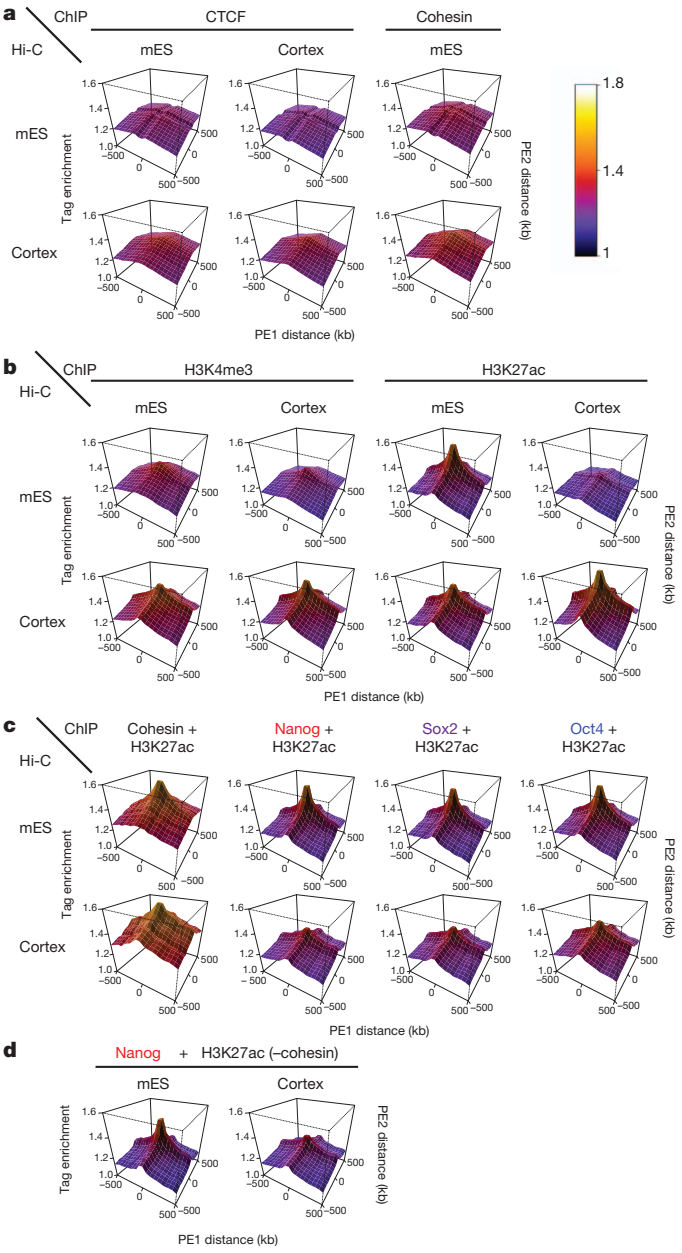


Figure 3. Spatial interactome of chromatin factors is revealed by PE-SCAn. A-D) PE-SCAn plots for various chromatin factors. Plots are the same as in Fig. 2e, but with a different scale on the vertical axis. Note that the height of the data is color coded according to the color bar shown in figure A). The top row represents the mESC Hi-C data, whereas the bottom row represents the cortex Hi-C data. A) PE-SCAn plots for known looping factors CTCF and cohesin (Smc1). B) PE-SCAn plots for promoter (H3K4me3) and active enhancer (H3K27ac) marks in mESC and cortex. C) PE-SCAn plots for active ▶

demonstrating that Nanog plays a direct role in bringing together distantly located clusters of Nanog binding sites.

Our data show that pluripotent transcription factors shape the pluripotent genome via spatial intra- and interchromosomal gathering of high-density binding sites. It has been suggested previously that transcription factors position tissue-specific and co-regulated genes in somatic cells[20-22]. However, in contrast to previous studies, we validated this concept by comparing genome-wide contact maps generated in wild type and transcription factor knockout cells and by studying an artificially induced cluster of binding sites. Our observation that targeting or removing a given factor to or from the genome only changes specific contacts while the overall folding of chromosomes remains intact is in accordance with a recently proposed model for chromosome topology. This 'dog-on-a-lead' model predicts that chromosomes are dominant over their individual segments (genes, domains, enhancers) in dictating the overall shape of the genome, but that segments can search the nuclear subvolumes they occupy for preferred contact partners [23]. There is accumulating evidence that stochastically determined nuclear environments can influence the transcriptional output of resident genes, leading to cell-to-cell variability[24, 25] We hypothesize that the observed spatial clustering of pluripotency factor binding sites in PSCs enhances the transcription efficiency of nearby genes and thereby contributes to the robustness of the pluripotent state.

METHODS SUMMARY

4C: sequencing & mapping

4C-sequencing was performed as previously described [9]. We used HindIII as the first restriction enzyme to generate the 3C template, which was further trimmed with DpnII. Sequencing was performed on Illumina GAII and HiSeq 2000 over multiple runs. Raw sequencing data and mapped wig files can be found under GEO accession GSE37275.

PE-SCAn

To assess which factors are associated with genome organization, we aligned ChIP data to the Hi-C data. For this, intrachromosomal captures that are >5 Mb from each other are aligned to transcription factor binding sites. Only captures where both di-tags mapped within 500kb of a ChIP peak were considered in the analysis. As a result we get for every pair of ChIP peaks on the same chromosome a set of two distances (*dx*, *dy*), to all the Hi-C di-tags that are found within 500kb of these peaks. From the distribution of *dx* and *dy* a frequency matrix is calculated with a bin size of 50 kb, which is normalized by dividing by a randomized dataset that is calculated by aligning the Hi-C data to a circularly permuted ChIPseq dataset, i.e. the ChIP peaks are linearly shifted 10 Mb along the chromosome. In

▶ enhancer sites co-bound by cohesin (Smc1), Nanog, Sox2 or Oct4. D) PE-SCAn plots for genomic sites with active enhancer marks and Nanog binding, but which are devoid of cohesin. Note that the left plot shows mESC Hi-C data whereas the right plot shows cortex Hi-C data.

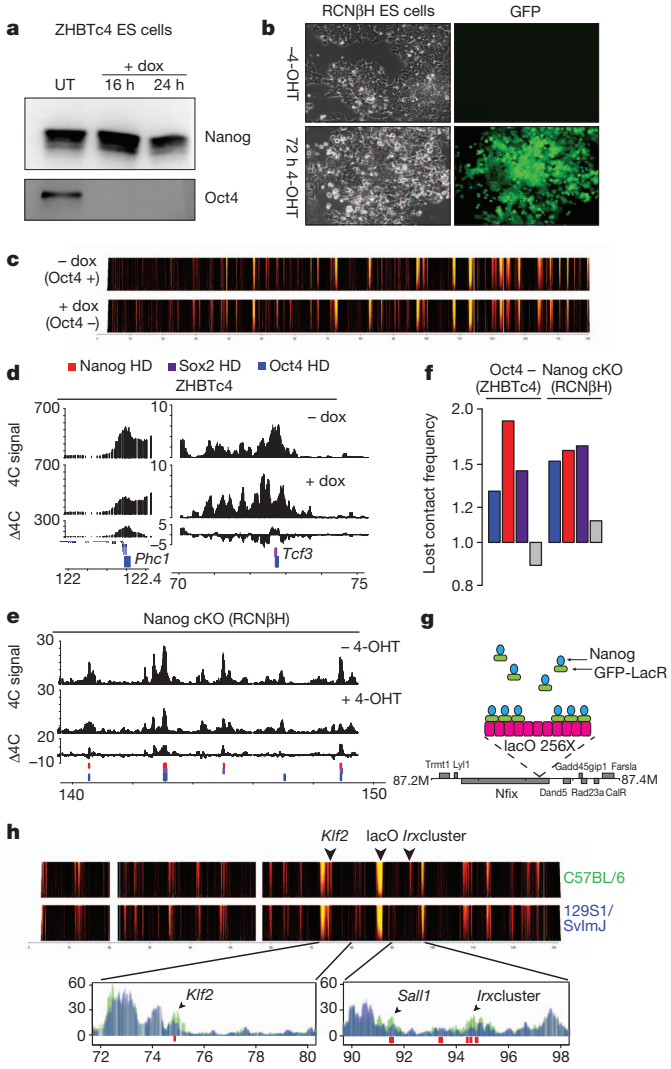


Figure 4. Pluripotency factors influence the 3D organization of the genome A) Immunoblot analysis before and after treatment of ZHBTc4 cells with 1 ug/ml doxycycline for the indicated times. Oct4 and Nanog proteins were detected using anti-Oct4 and anti-Nanog antibody. B) Morphology and GFP expression of RCNβH cells before and after 72h treatment with 1uM tamoxifen. C) 4C domainograms for Oct4 positive (-dox) and Oct4 negative (+dox) ZHBTc4 cells show that overall chromosome topology is maintained in Oct4 depleted cells. D) Zoomed in regions show 4C signal (reads/million) for Oct4 positive cells (top panels) and Oct4 negative cells (middle panels). Bottom panel shows the difference (delta) between the 4C signal of the Oct4 positive and negative cells. Red, purple and blue rectangles show high-density Nanog, Sox2, and Oct4 regions, respectively. E) Same as D) but for Nanog conditional knock-out. Note that gene information is left out at this scale. F) Chromosome-wide analysis of differential 4C interactions for the *Nanog* enhancer viewpoint. Loss of 4C contact frequency is defined as a lower 4C signal in the knock-out compared to the non-depleted reference cell ▶

this manner the structure of the Hi-C data is preserved; the structure of the ChIP data is also preserved, only shifted.

Depletion of pluripotency factors

RCNβH cells were treated with tamoxifen and replated the following day. 72h after initial tamoxifen treatment, cells were harvested for 4C template preparation and analyses. ZHBTc4 cells were harvested after 48h of treatment with 1 μg/ml doxycycline.

ACKNOWLEDGEMENTS

We would like to thank Carlo Vermeulen and Sjoerd Holwerda for counting FISH slides; Geert Geeven for the analysis of sequencing data; Johan Brandsma (Erasmus MC) for technical assistance; Dr. Pernette Verschure (Swammerdam Institute for Life Sciences, the Netherlands) for the LacR-GFP backbone construct; and Dr. Hitoshi Niwa (RIKEN Center for Developmental Biology, Japan) for kindly providing us with ZHBTc4 ES cells. We also thank the Netherlands Institute for Regenerative Medicine (NIRM) network for supporting the RAP laboratory and the Medical Research Council of the UK for supporting the Chambers laboratory.

This work was financially supported by grants from the Dutch Scientific Organization (NWO) to EdW (700.10.402, 'Veni') and to WdL (91204082 and 935170621), InteGeR FP7 Marie Curie ITN (PITN-GA-2007-214902) and a European Research Council Starting Grant (209700, '4C') to WdL.

AUTHOR CONTRIBUTIONS

EdW conceived the study, analyzed the data and wrote manuscript. BB designed and performed reprogramming and knockout experiments, and helped writing the manuscript. YZ and PHK designed and performed LacR-Nanog experiments. ES and PK performed cell culture and 4C experiments. MV, EN and EH designed, performed and analyzed FISH experiments. MW and NG assisted with reprogramming experiments. RP shared Oct4 conditional knockout cells and assisted with depletion experiments. NF and IC shared conditional knockout cells and assisted with Nanog depletion experiments. WdL conceived the study and wrote manuscript

► line. Loss of contact frequency is determined for high-density Oct4, Nanog, Sox2 and CTCF (control) clusters for the Oct4 ablated and Nanog conditional knock-out cell lines and the enrichment over the background is calculated (see Methods for details). G) Schematic drawing depicting the integration site of the lacO repeat cassette in the C57Bl/6 allele of the *Nfix* locus and the targeting of GFP-LacR-Nanog fusion proteins. H) Domainograms showing allele-specific 4C[19] for the C56BL/6 (harboring the lacO cassette) and the 129 allele. Lower panels show zoomed-in 4C profiles (C57Bl/6 green, 129 blue) for example differential regions. Red rectangles indicate high-density Nanog clusters (6 sites / 100kb).

AUTHOR INFORMATION

4C data has been submitted to GEO under accession number GSE37275: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37275 Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interest. Correspondence and requests for materials should be addressed to w.delaat@hubrecht.eu.

REFERENCES

1. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-80 (2012).
2. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-5 (2012).
3. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. Genes Dev 26, 11-24 (2012).
4. van Steensel, B. & Dekker, J. Genomics tools for unraveling chromosome architecture. Nat Biotechnol 28, 1089-1095 (2010).
5. Splinter, E. & de Laat, W. The complex transcription regulatory landscape of our genome: control in three dimensions. EMBO J 30, 4345-55 (2011).
6. Simonis, M. et al. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). Nat Genet 38, 1348-54 (2006).
7. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289-93 (2009).
8. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. Science 295, 1306-11 (2002).
9. Splinter, E. et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev 25, 1371-83 (2011).
10. 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, 221-30 (2012).
11. Wutz, A., Rasmussen, T. P. & Jaenisch, R. Chromosomal silencing and localization are mediated by different domains of Xist RNA. Nat Genet 30, 167-74 (2002).
12. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28, 495-501 (2010).
13. Handoko, L. et al. CTCF-mediated functional chromatin interactome in pluripotent cells. Nat Genet 43, 630-8 (2011).
14. Lin, Y. C. et al. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. Nat Immunol 13, 1196-204 (2012).
15. Li, G. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148, 84-98 (2012).
16. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116-20 (2012).
17. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24, 372-6 (2000).
18. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230-4 (2007).
19. Holwerda, S. J. et al. Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. Nucleic Acids Res (2013).

20. Schoenfelder, S. et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. Nat Genet 42, 53-61 (2010).
21. Xu, M. & Cook, P. R. Similar active genes cluster in specialized transcription factories. J Cell Biol 181, 615-23 (2008).
22. Dhar, S. S. & Wong-Riley, M. T. Chromosome conformation capture of transcriptional interactions between cytochrome c oxidase genes and genes of glutamatergic synaptic transmission in neurons. J Neurochem 115, 676-83 (2010).
23. Krijger, P. H. & de Laat, W. Identical cells with different 3D genomes; cause and consequences? Curr Opin Genet Dev (2013).
24. Noordermeer, D. et al. Variegated gene expression caused by cell-specific long-range DNA interactions. Nat Cell Biol 13, 944-51 (2011).
25. Kind, J. et al. Single-cell dynamics of genome-nuclear lamina interactions. Cell 153, 178-92 (2013).
26. Feng, B. et al. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol 11, 197-203 (2009).
27. Karantzali, E. et al. Sall1 regulates embryonic stem cell differentiation in association with nanog. J Biol Chem 286, 1037-45 (2011).
28. Parisi, S. et al. Klf5 is involved in self-renewal of mouse embryonic stem cells. J Cell Sci 121, 2629-34 (2008).
29. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553-60 (2007).
30. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766-70 (2008).

ONLINE-ONLY METHODS

Cell culture

E14 ESCs (129/Ola background) and C56Bl/6-129 ESCs were grown in BRL conditioned DMEM (high glucose, Gibco) supplemented with 15% FBS, 1X NEAA (Gibco), 1X Pen Strep (Gibco), 1:1000 β-mercaptoethanol (Invitrogen), 1X L-Glutamine (Gibco) and 1000 U/ml LIF (Gibco). Independently derived 129/Cast ESCs (129SVJ/Castaneus background) were grown on irradiated mouse embryonic fibroblasts (MEFs) in DMEM supplemented with 15% FBS, 1x NEAA, 1x Pen Strep, 1:1000 β-mercaptoethanol and 1000 U/ml LIF. RCNβH cells were cultured in GMEM/β-mercaptoethanol/10%FCS/LIF as described previously [31] and [32]. ZHBTc4 [33] cells were cultured in GMEM/β-mercaptoethanol/15%FCS/Sodium Bicarbonate/ LIF. Culture media were supplemented with 1ug/ml doxycycline or 1 uM tetracycline when applicable. NP cells (E14 and 129/Cast) were grown in DMEM-F12 supplemented with 1:100 N2 (Gibco), 20 ng/ml bFGF (Peprotech), 20 ng/ml murine EGF (Peprotech). For the 129/Cast NP cells 1X B-27 (Gibco) was added [34]. We generated astrocytes by growing E14NP cells to confluency and washing twice with DMEM before adding astrocyte medium (DMEM-F12 supplemented with 1:100 N2 and 2% FBS) [35]. The culture medium was changed twice and cells were grown for 5 days to make sure differentiation was complete, which was confirmed by immunofluorescence.

Generation of induced pluripotent stem (iPS) cells

For generation of iPS cells, 10.000 129/Cast neural progenitor (NP) cells were seeded on gelatin-coated dishes in N2B27 medium (StemCell Resources). Cells were infected o/n with lentivirus expressing a multicistronic reprogramming cassette, encoding the iPS factors Oct4, Klf4, Sox2 and c-Myc [36]. After 5 days, cells were harvested and plated on irradiated MEFs. On day 6, N2B27 medium was replaced with mouse ES cell medium (DMEM with 15% FBS, 1x NEAA, 1x Pen Strep, 1:1000 β-mercaptoethanol and 1000 U/ml LIF). iPS colonies were picked for clonal expansion on day 20-22 after infection. At passage 11 after colony picking, proper iPS reprogramming was examined by qPCR analysis on a panel of marker genes on total RNA (pluripotency markers: Nanog, Zic3, Dppa4, Sall4, Cer1, Sox17 and Fgf5, neuronal lineage markers: Olig2, Nestin, Blbp and Glast). Cells were harvested for 4C at passage 11.

siRNA knockdown of Nanog

For our knockdown experiments we used a pool of siRNA oligos targeting *Nanog* (M-057004-01) and a control pool containing non-targeting siRNAs (D-001206-13, siGENOME SMARTpool, Dharmacon). 129/Cast ESCs were seeded without feeders in 100mm culture dishes at ~20% confluency on the day prior to transfection. Cells were transfected according to the manufacturer's protocol using 25nM final siRNA concentration combined with 50 ul DharmaFECT 1. Transfection mixtures were added directly into the culture medium and plates were incubated at 37°C with 5% CO_2. 48 h after transfection, cells were harvested for protein level analysis and 4C template preparation.

Conditional ablation of Nanog and Oct4

RCNβH cells were treated with tamoxifen and replated the following day. 72h after initial tamoxifen treatment, cells were harvested for 4C template preparation and analyses. ZHBTc4 cells were harvested after 48h of treatment with 1 ug/ml doxycycline.

Protein analysis

Protein levels before and after conditional deletion were analyzed in cells harvested at the time points as described above. Immunoblot analysis was carried out on nuclear extracts that were made as described in [37]. Extracts were subjected to sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) [38], and proteins were transferred to a methanol-activated PVDF membrane. Blots were blocked in blocking buffer (5% non-fat dry milk in TBST (50mM Tris pH 7.4/150mM NaCl/0.1%Tween)) for 1h at room temperature or overnight at 4°C, while tumbling. Primary antibody was diluted in blocking buffer and incubated for 1-3h at room temperature or overnight at 4°C, while tumbling. Blots were washed four times with TBST and incubated with secondary antibody for 1h in blocking buffer. Membranes were then incubated with SuperSignal West Pro (Thermo Scientific) and digitally analyzed using an LAS 4000 ECL ImageQuant imager and ImageJ software. Used antibodies: anti-Nanog (A300-397A, Bethyl Laboratories Inc.) at 1:5000, anti-Oct4 (C30A3, Cell Signaling Technology Inc) at 1:1000, anti-histone H3 (Abcam 1791) at 1:2000.

Flow Cytometry

Tamoxifen treated and untreated RCNβH cells were trypsinized and pellets were resuspended as single cells in regular ESC medium at about 10^6 per ml. For each condition, 50000 live cells were analyzed for GFP fluorescence, using a Becton-Dickinson FACSCalibur flow cytometer and FloJo software.

Generation of lacO targeted cell line

Homology arms were excised (KpnI digest) from BAC RP24-136A15, and ligated into a low copy bluescript plasmid. Two hundred and fifty-six copies of a LacO array were inserted into a unique AatII site of the homology arms. F1 ES cells derived from C56Bl/6 and 129 mouse strains were transfected with the linearized targeting construct by electroporation. After 14 days of selection with neomycine, positive colonies were picked and screened by Southern blotting. The GFP-LacR-Nanog construct was generated in the backbone of pHAGE2-IRES-puro with an EF1α promoter [39]. LacO cells were stably transduced with the GFP-LacR-Nanog construct, and positive cells were selected with puromycine for 10 days after which cells were harvested and tested for purity of by flow cytometry (90% GFP-positive). Allelic paired-end 4C technology was performed as described [40], using HindIII-DpnII digestion and the following 4C primers: AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCC GATCGGAACTAAATGGAGGATC and CAAGCAGAAGACGGCATACGAGATCGGTCT CGGCATTCCTGCTGAACCGCTCTTCCGATCTTACCAGGACCCCTGGGACCC

3D DNA FISH

3D DNA-FISH for interchromosomal interaction was performed essentially as described in [41]. **Slide preparation.** ESCs were spotted on polylysine microscopy slides after which

slides were washed in PBS. Cells were fixed in 3% paraformaldehyde/PBS and washed twice with PBS, after which cells were permeabilized on ice using ice-cold 0.5% TritonX100 for 6 min. Slides were then washed for 3 minutes with 70% Ethanol and stored in 70% Ethanol at -20°C. **Probe preparation.** For preparation of probes, 10ul of both labeled BACs was combined with 5ul mouse Cot1 DNA and mixtures were speedvacced until pellets were dry. Pellets were resuspended in 12.5ul 50+ hybmix, incubated for 5 min at 95°C, cooled on ice, and incubated for 30 min at 37°C. **FISH hybridization.** Slides were dehydrated for 3 min in 70% Ethanol, 3 min in 90% Ethanol, 3 min in 100% Ethanol, after which slides were air-dried. 100ul 70+ hybmix was then added to the dried slides, and slides were covered with a coverslip and incubated for 3 min at 85°C. Slides were washed on ice, using ice-cold 2x saline-sodium citrate buffer (SSC) for 5 min, then using ice cold 70% Ethanol for 5 minutes, after which slides were dehydrated again as described above. After air drying, 10ul probe was added and covered with a coverslip and hybridizing slides were incubated overnight at 37°C in a humid box containing 50% formamide/2xSSC. **Post hybridization**. Slides were washed in 2X SSC for 5 minutes, which also removes the coverslip. Subsequently, slides were washed 3 times for 10 minutes in 50% formamide/2xSSC at 37°C. Slides were then dehydrated as described above, and air dried slides were mounted using 40ul DAPI/Vectashield. Slides were covered with new coverslips and sealed with transparent nail polish. We performed manual distance measurements in ImageJ using the Image5D plugin.

General 4C template information

For high quality 4C experiments library complexity is crucial; by applying 4C to 1 million genome equivalents (3 µg DNA), we analyze a large number of ligation products per viewpoint. The generated DNA contact profiles are therefore a true population average[42]. The observed ligation products are the result of spatial proximity. Note that these ligation products can be a reflection of direct DNA contacts (such as promoter-enhancer interactions) or indirect contacts mediated by large macromolecular complexes or nuclear particles.

Experimental and primer design is done as previously described[43]. For the allele-specific 4C we have used a paired-end 4C strategy[40]. To this end, we designed forward and reverse primers compatible with the Illumina flow cell. The forward primer analyzes the ligation product and the reverse primer is selected such that it sequences a SNP that distinguishes the C57Bl/6 allele from the 129S1/SvImJ allele. After sequencing this SNP is used to demultiplex the two alleles, to create two separate 4C profiles.

Definitions

To make this methods section clearer to non-experts we present the following definitions. **Fragment**: A genomic region (or sequence) that is generated after the first restriction. In this case, the first restriction enzyme, that generates the 3C template, is always HindIII. **Fragment end**: To generate the 4C template, the 3C template is further digested with a frequent cutter, in our case DpnII. The resulting HindIII-DpnII restriction fragment is referred to as the fragment end, because this restriction fragment represents the end of the 3C fragment. **Capture frequency**: Captures are defined as the ligations in the 3C protocol

resulting from 3D genome conformation. The 4C primers directly interrogate the ligation junction. Therefore the resulting capture frequency can be estimated from counting the number of reads coming from a given fragment end.

4C: sequencing & mapping

4C-sequencing was performed as previously described [34]. We used HindIII as the first restriction enzyme to generate the 3C template, which was further trimmed with DpnII. Sequencing was performed on Illumina GAII and HiSeq 2000 over multiple runs. Primer sequence (internal barcode) is removed from the sequence and the trimmed reads were aligned to a reduced genome consisting of sequences that flank HindIII restriction sites. The mouse mm9 genome was used as the reference genome for mapping 4C sequence captures. Non-unique sequences (repeats) that flank a restriction site were removed from the analysis. From the mapping a frequency distribution along the genome is calculated, which is the input for all downstream analyses. Raw sequencing data and mapped wig files can be found under GEO accession GSE37275.

Statistical analysis of 4C data

Statistical analyses of 4C data (i.e. domainograms and target identification) was performed as described in refs [34, 43]. For formal definitions we refer the reader to these articles. Here we will briefly describe underlying principles of the data analysis. An inherent challenge of 4C data (and genome-wide chromosome capture data in general) is the highly non-uniform data distribution. Close to the viewpoint the signal is very high, whereas the signal rapidly decreases as a function of the distance from the viewpoint. Therefore, we statistically define significant interactions as regions that have an increased number of captures compared to the *local* background. To this end we must estimate the local background capture frequency. To minimize potential PCR artifacts we transform the 4C-seq read count at HindIII-DpnII fragment ends to binary data (i.e. captured or not captured). From this it is clear that 0s play an essential role in determining significant interactions. Local background is then determined as the frequency of captured fragment ends (1s) in a large window, typically 3000 fragment ends. Following the binomial distribution, we can estimate μ and σ (for details see [43], which is used to determine a z-score for a window of fragment ends of fixed size.

In order to visualize the 4C data using domainograms z-scores are calculated using windows with a range of size (from 3 through 200). The z-scores are subsequently transformed to p-values with a one-tailed normal test. The $-\log_{10}$-tranformed p-values are color coded and visualized along the linear chromosome. As such regions can be visualized with a high likelihood of interaction with the viewpoint.

In order to distill discrete regions of significant interaction we choose a fixed window size of 100 fragment ends and calculate the z-scores for this window size across the chromosome. To select significant regions we determine the z-score threshold based on a FDR level of 0.01. The FDR is determined based on the z-score distribution in 100 randomly permuted chromosomes. The windows exceeding the z-score threshold are selected as significantly contacted regions.

Analysis of 4C trans interactions

In our dataset we find highly specific interchromosomal interactions. Like the intrachromosomal profiles we calculate an enrichment score over the background capture frequency. However, because the background capture frequency is distributed more or less uniformly across the chromosome, we can use a single background frequency per chromosome. The 4C enrichment score is calculated in the following way:

$$E_{\text{trans},i} = \frac{w \cdot p_{w,i} - (p_{\text{chrom}} \cdot w)}{\sqrt{w \cdot p_{\text{chrom}} \cdot (1 - p_{\text{chrom}})}} \quad (1)$$

where w is the window size, and i is window index along the chromosome. p_{chrom} is defined as follows:

$$p_{\text{chrom}} = \frac{N_{\text{captured}}}{N_{\text{chrom}}} \quad (2)$$

Where N_{captured} is the number of fragment ends captured on the chromosome and N_{chrom} is the total number fragment ends on the chromosome. $p_{w,i}$ is defined as follows:

$$p_{w,i} = \frac{n_{w,i,\text{captured}}}{w} \quad (3)$$

Where $n_{w,i,captured}$ is the number of fragments captured in genomic window i of size w. Windows with an $E_{\text{trans,I}}$ larger than 6 were chosen for subsequent analysis in GREAT.

4C/Hi-C alignment to ChIP profiles

To test enrichment of 4C signal along ChIP peaks we aligned the *trans* fragments to nearest ChIP peak. We used several ES cell ChIP-seq profiles from various sources. For Oct4, Sox2 and Nanog we used the data from Marson et al [44] (GSE11724). H3K27ac was taken from Creyghton et al. [45] (GSE24165). Smc1 data [46] (GSE22557), H3K4me3 [47] (GSE12241), RNA PolII and CTCF (mESC and cortex), H3K27ac and H3K4me3 (cortex) [48] (GSE29218).

4C data was binarized because in trans the capture frequency is so low that read count more likely represent differences in PCR efficiency rather than genuine unique captures. For the binarized data the distance to the nearest ChIP peak was calculated. In order to calculate enrichment scores, the distances to the nearest ChIP peak were sorted (i.e. aligned) and a sliding average was calculated. The window size of the sliding average was set to 1% of the total dataset.

Hi-C normalization and analysis

HiC data [49] was downloaded from GEO (accession GSE35156). We removed all read pairs that are mapped within 500 bp of each other on the chromosome, because these read pairs are likely genomic background sequence, rather then *bona fide* HiC captures. For the virtual 4C and disorganization analysis we average the data to bins of 100kb, which results in a matrix of pairwise capture frequencies between all the genomic bins. A proper analysis of the HiC data requires that we to correct the HiC matrix for genomic biases. For this normalization

we assume that the capture probability in a given genomic bin is dependent on the number of restriction sites in this bin. The strong positive correlation between the restriction site density and the number of captures for that given bin is evidence that this assumption is correct (data not shown). We therefore normalize the bins by dividing by the capture probability. First we calculate the restriction density in 100 kb bins along the chromosome, which gives us a capture probability for in a given bin ($p_{capture,i}$). The capture probability between two bins on the chromosome ($P_{capture,i,j}$) can now be calculated by taken the product of the capture probabilities of the two single bins ($p_{capture,i}$, $p_{capture,j}$). Before normalization the correlation between the diagonals of the the HiC matrices for the NcoI and HindIII experiments from mouse ES cells is 0.32. However, after normalization this correlation has jumped to 0.86.

For the virtual 4C based on the HiC data, we combine the three normalized matrices (2x HindIII, 1x NcoI). Because the data is too sparse to perform a virtual 4C analysis for a single fragment we analyze a single row from the HiC interaction matrix. For comparison, we also calculate the average 4C signal in 100kb genomic bins.

For the analysis of genomic disorganization we use the two HindIII experiments for mESC and cortex (GSE35156). In this analysis we want to compare the propensity of active and inactive regions to contact regions over large genomic distances. To this end we segment the chromosomes in active and inactive bins of 100kb, based on the density of H3K4me1 sites[50]. Based on this segmentation we can create a matrix with similar in size to the HiC matrix. In this matrix three classes of interaction bins can be created:

1. H3K4me1 high in both: interaction bin between two active genomic regions
2. H3K4me1 low in both: interaction bin between two inactive genomic regions
3. H3K4me1 low / H3K4me1 high: interaction bin between and active and inactive

Because we perform a 50/50 segmentation, the classes H3K4me1 high/high and H3K4me1 low/low will both be 25% of the interaction bins, H3K4me1 high/low class will be 50% of the interaction bins. In addition, the HiC matrix is segmented into high-contact and low-contact bins by setting an arbitrary threshold (75% quantile value of the entire matrix). Next, we overlay the segmented HiC matrix and the contact bins to determine the number of long-range contacts made for each of the classes. We use various minimal distance cut-offs running from 10 Mb to 70Mb with step sizes of 10 Mb. This process is schematically explained in Figure 1e.

Alignment of Hi-C data to ChIP peaks (PE-SCAn)

To assess which factors are associated with genome organization, we aligned ChIP data to the HiC data. To this end we selected the intrachromosomal captures, however, because of the strongly non-uniform distribution we removed the captures that lie within 5 Mb of each other. This has the effect that we only analyze interactions between, rather than within, topological domains. The HiC pairs were aligned to the ChIP data in two iterations. First, one end of the paired reads was aligned to the ChIP data. Only reads that mapped within 500 kb up- or downstream of the ChIP peaks were selected for further analysis. Of this reduced set the corresponding read was also aligned to the ChIP peaks within 500kb. As a result we

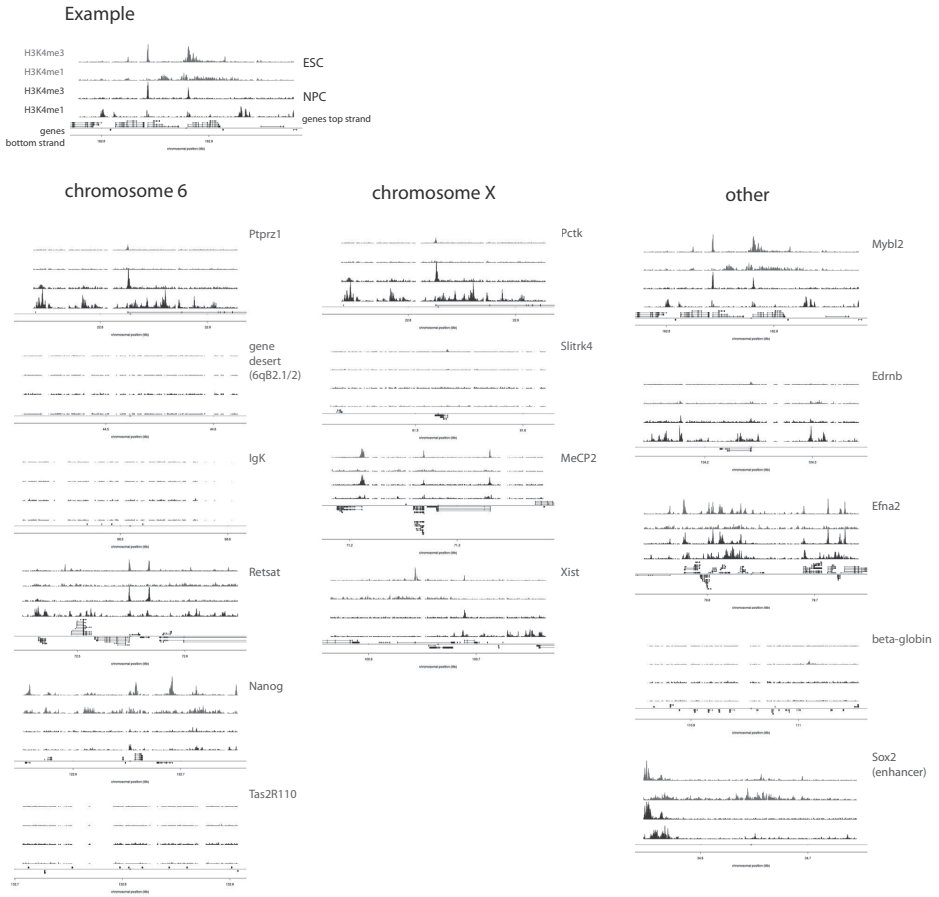
get for every intrachromosomal pair of ChIP peak a set of two distances (*dx*, *dy*), to all the Hi-C ditags that are found within 500kb of these peaks. From the distribution of *dx* and *dy* a frequency matrix is calculated, which is the result of our two-dimensional alignment, with a bin size of 50 kb. In order to calculate whether the binding sites of a given factor show preferential spatial contacts we calculate an enrichment score over a randomized dataset. The randomized dataset is calculated by aligning the Hi-C data to a circularly permuted ChIPseq dataset, i.e. the ChIP peaks are linearly shifted10 Mb along the chromosome. It is important to note that in this manner the structure of the Hi-C data is preserved; the structure of the ChIP data is also preserved, only shifted. The resulting frequency matrix serves as an internal normalization for the observed Hi-C data alignment scores.

REFERENCES

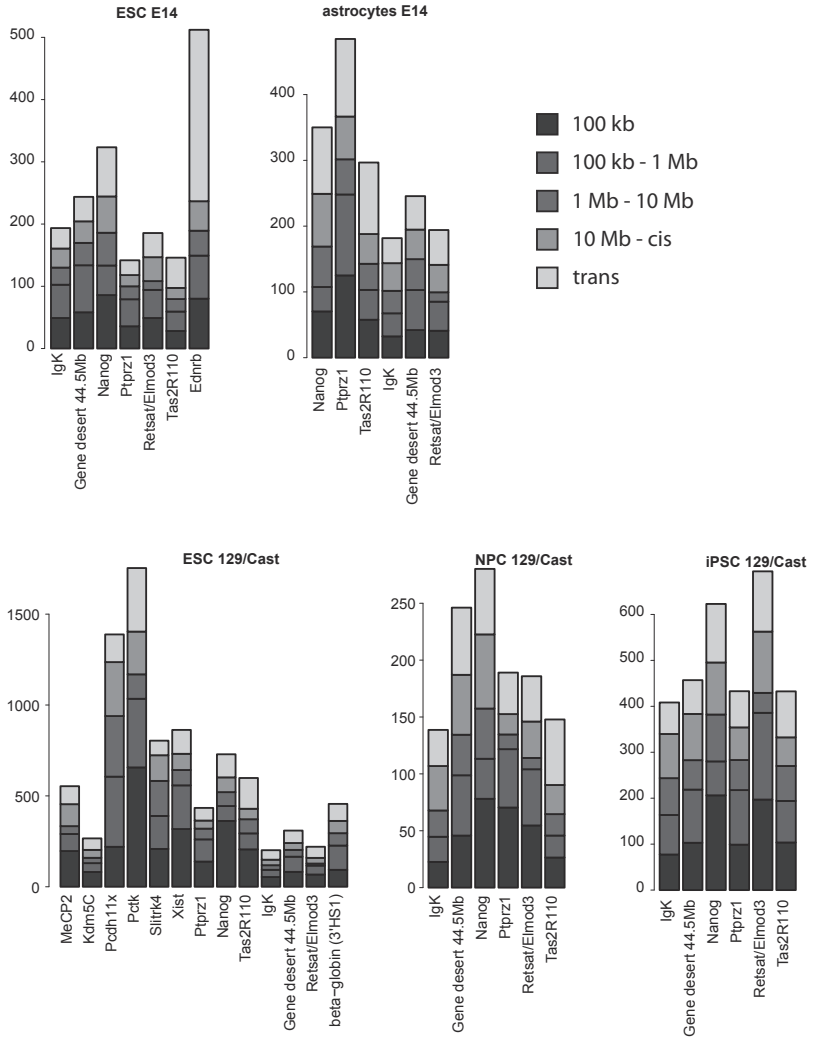
31. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230-4 (2007).
32. Smith, A. G. Culture and differentiation of embryonic stem cells. J Tissue Cult Methods 13, 89-94 (1991).
33. Niwa, H., Miyazaki, J. & Smith, A. G. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet 24, 372-6 (2000).
34. Splinter, E. et al. The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev 25, 1371-83 (2011).
35. Peric-Hupkes, D. et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 38, 603-13 (2010).
36. Warlich, E. et al. Lentiviral vector design and imaging approaches to visualize the early stages of cellular reprogramming. Mol Ther 19, 782-9 (2011).
37. Andrews, N. C. & Faller, D. V. A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res 19, 2499 (1991).
38. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-5 (1970).
39. Wilson, A. A. et al. Sustained expression of alpha1-antitrypsin after transplantation of manipulated hematopoietic stem cells. Am J Respir Cell Mol Biol 39, 133-41 (2008).
40. Holwerda, S. J. et al. Allelic exclusion of the immunoglobulin heavy chain locus is independent of its nuclear localization in mature B cells. Nucleic Acids Res (2013).
41. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 485, 381-5 (2012).
42. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. Genes Dev 26, 11-24 (2012).
43. 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, 221-30 (2012).
44. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 134, 521-33 (2008).
45. Creyghton, M. P. et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A 107, 21931-6 (2011).
46. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. Nature 467, 430-5 (2010).

47. Mikkelsen, T. S. et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448, 553-60 (2007).
48. Shen, Y. et al. A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116-20 (2012).
49. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376-80 (2012).
50. Meissner, A. et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454, 766-70 (2008).

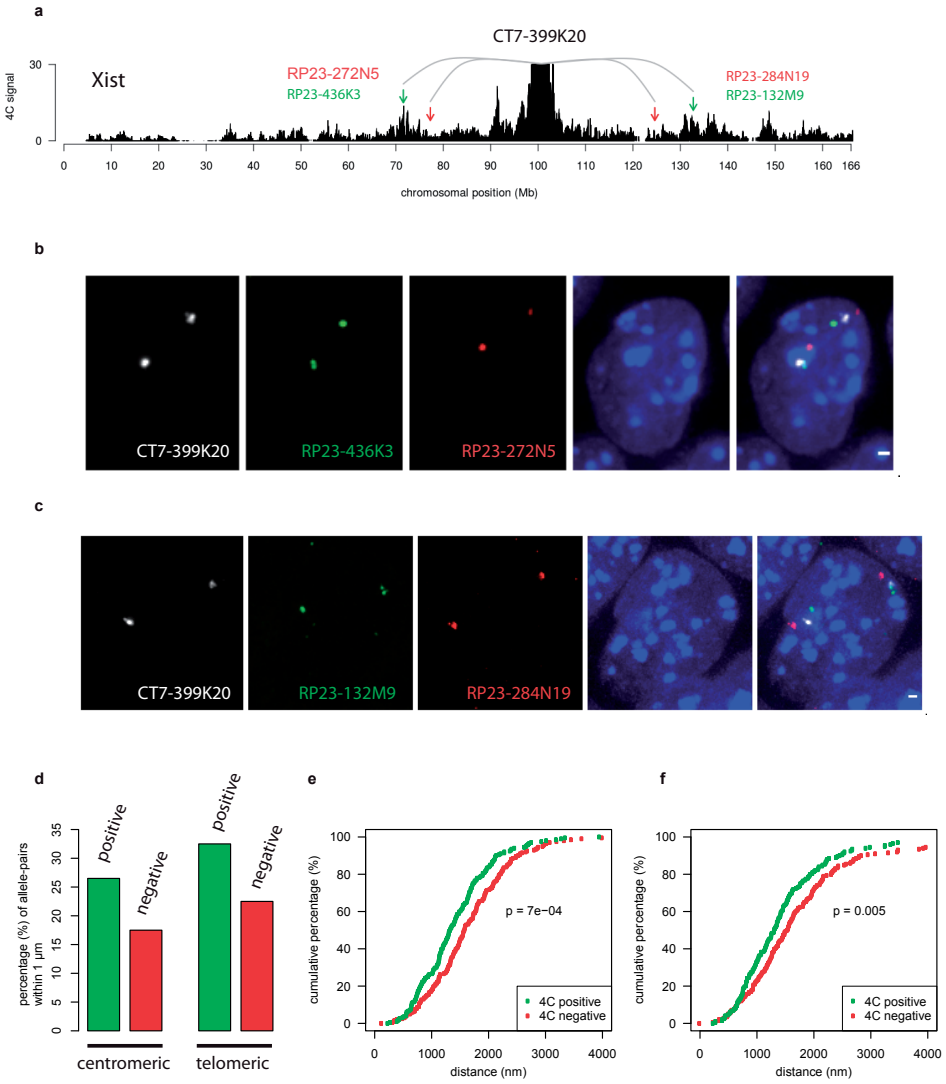
SUPPLEMENTARY MATERIALS



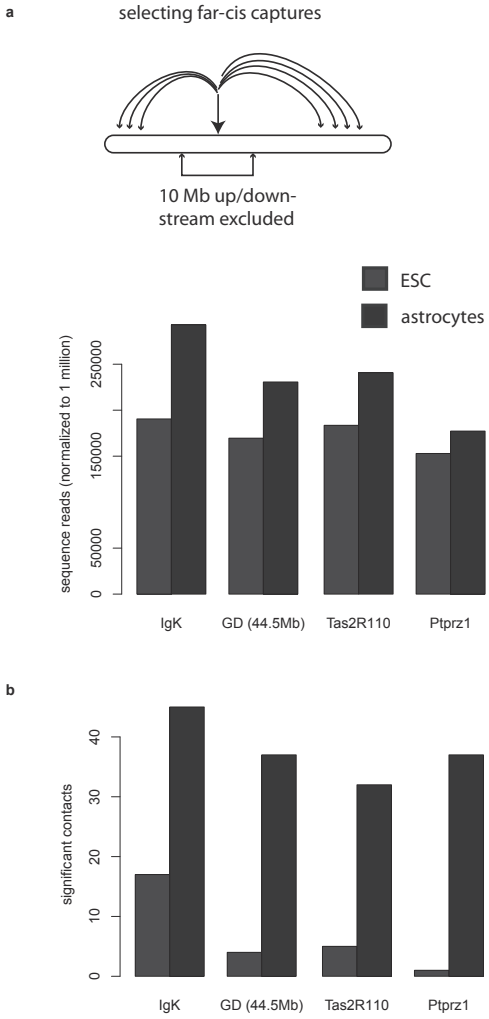
Supplementary Figure 1. Overview of viewpoints used in this study. Chromosomal context of the viewpoints is shown, with promoter (H3K4me3) and enhancer (H3K4me1) histone modification data in ESC and NPC [29, 30].



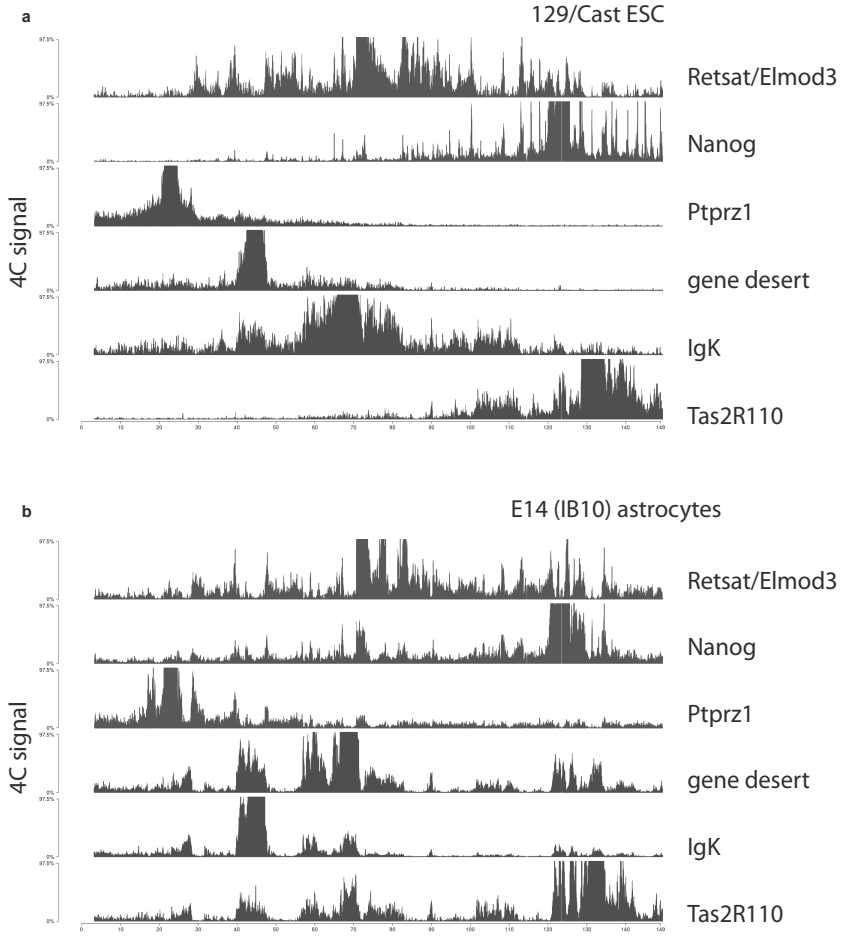
Supplementary Figure 2. Read distributions of chromosome 6 viewpoints in various tissues. Barplots summarize the read counts (x1000) for the various classes of captures. Captures are categorized in five classes depending on their genomic distance from the viewpoint: within 100kb (darkblue), 1 Mb (lightblue), 10 Mb (red) of the viewpoint, on the *cis* chromosome (orange), i.e. intrachromosomal captures and *trans* or interchromosomal captures in gray. This figure is based on the numbers in Supplementary Table 1.



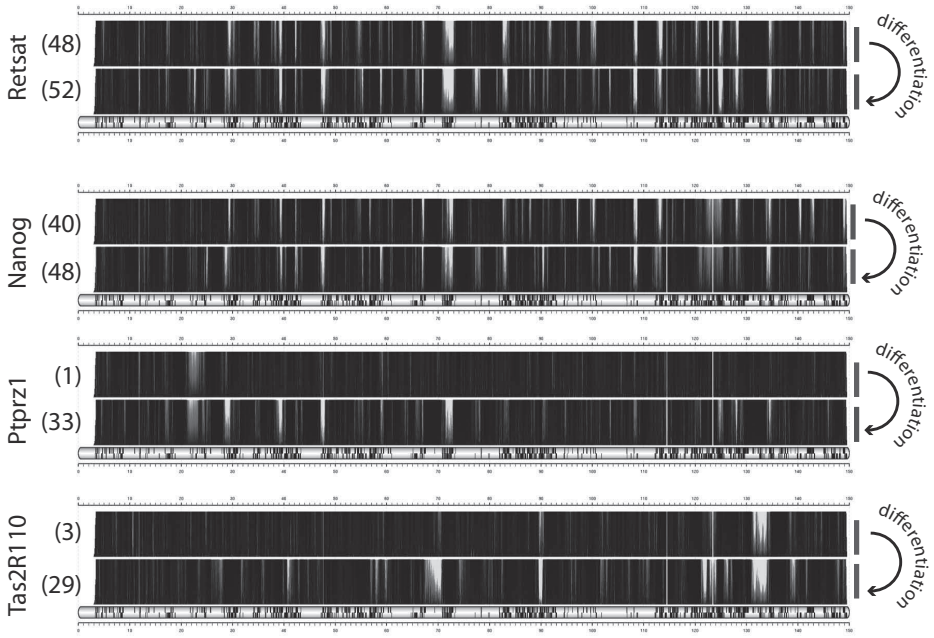
Supplementary Figure 3. 3D-DNA FISH confirms preferential contacts identified by 4C-seq. A) Contact profile for the X inactivation regulator *Xist*. Green arrows point to genomic regions identified as significantly contacting with *Xist*, red arrows point to regions not significantly contacted, but closer on the linear DNA template. B) Representative three color DNA-FISH images for the probes on the centromeric side of *Xist*. C) Representative three color DNA-FISH images for the probes on the telomeric side of *Xist*. D) Quantification of the number of spots within 1 µm, for the centromeric and telomeric probe combination. Cumulative distribution plot for spot distances for the centromeric E) and telomeric F) probe.



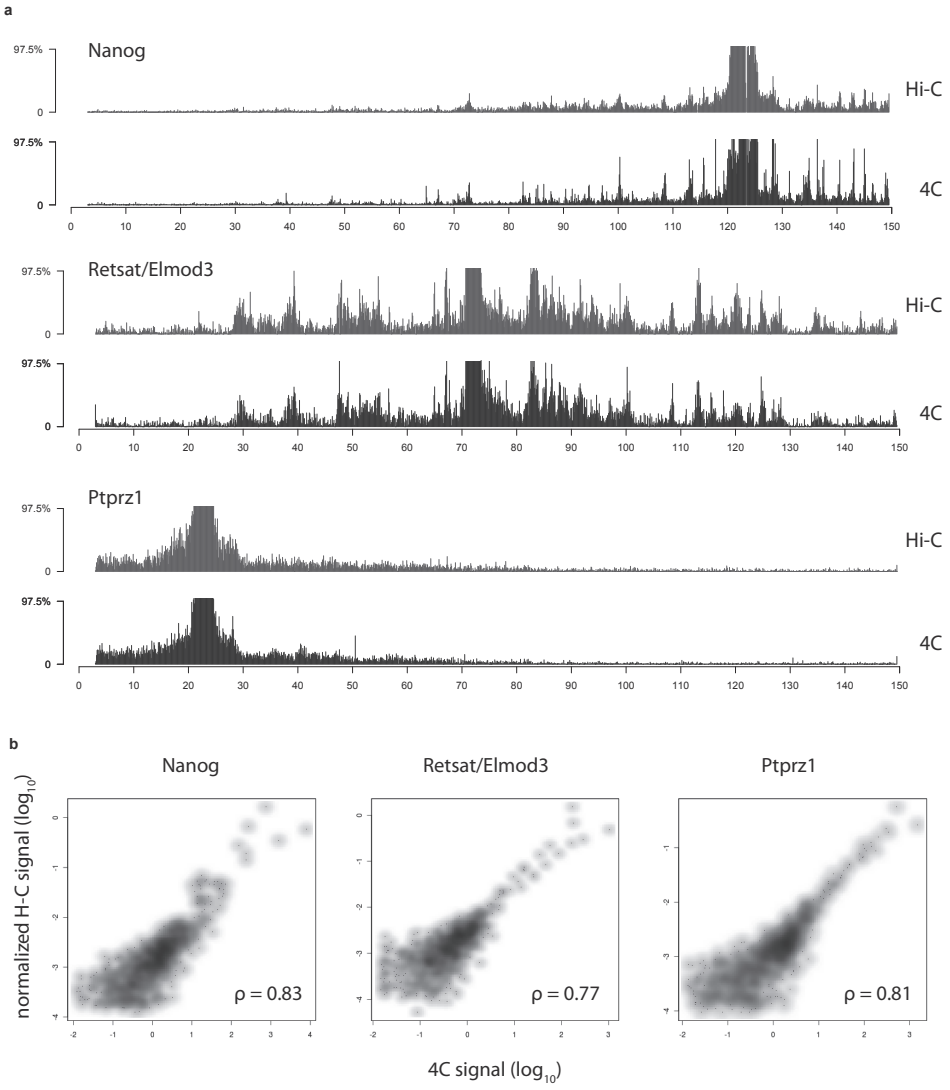
Supplementary Figure 4. Far-*cis* capture frequency for inactive viewpoints in ESC and astrocytes. A) The normalized number of intrachromosomal captures that are found more than 10 Mb from the viewpoint for E14 ES cells and astrocytes. B) Number of significantly contacted regions for same viewpoints and cell lines. The comparison shows that while each locus does slightly increase its balance between distal and local contacts upon differentiation (from 1,2-fold (Ptprz1) to 1,5-fold (Igκ)) this increase is not nearly as dramatic as the increase in number of specific contacts (which goes up from 1 to 33 regions for Ptprz1). Importantly, the ability of inactive regions to capture distal sequences in astrocytes is not different from that of any of the inactive regions in ESCs. This shows that the distribution of long-range captures, more than the capacity to reach over distance, is different and more random in ESCs.



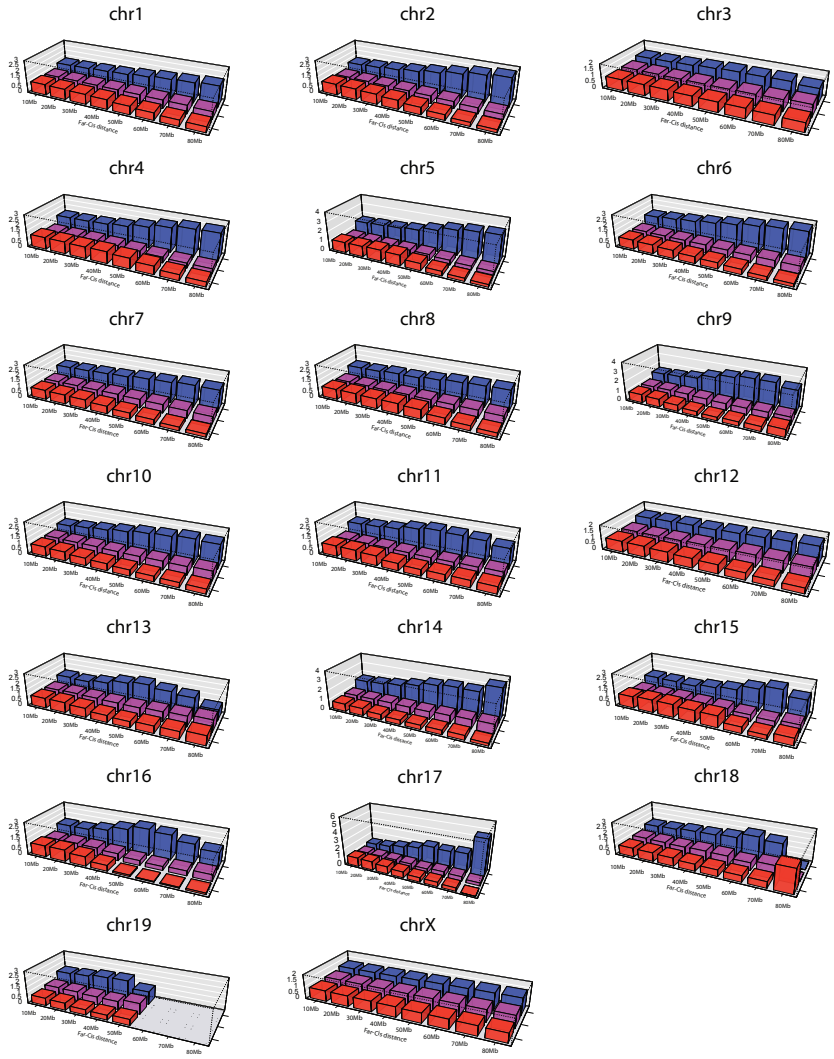
Supplementary Figure 5. Chromosomal maps of 129/Cast ESC and E14 astrocytes. Chromosomal maps for A) 129/Cast ESC and B) E14 astrocytes, calculated as in Figure 1A.



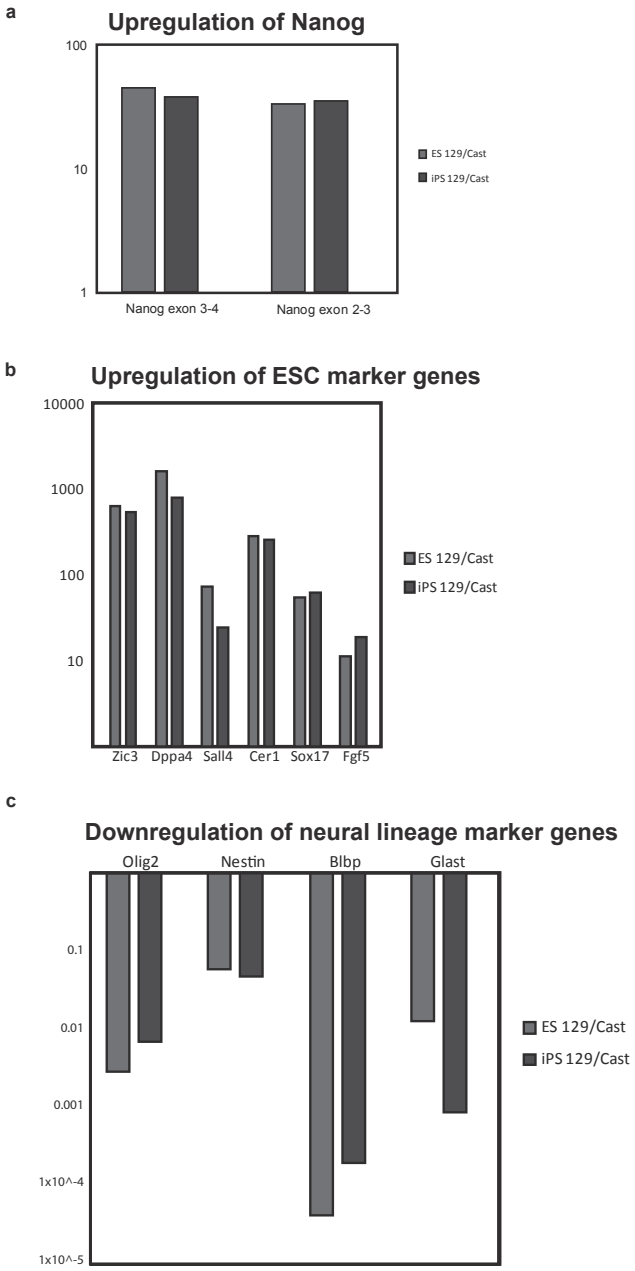
Supplementary Figure 6. Differentiation to astrocytes leads to more organized inactive chromatin. Domainograms show chromosomal topology of *Retsat*, *Nanog*, *Ptprz1* and *Tas2R110* in ESCs and astrocytes. Domainogram analysis as in Figure 1E..



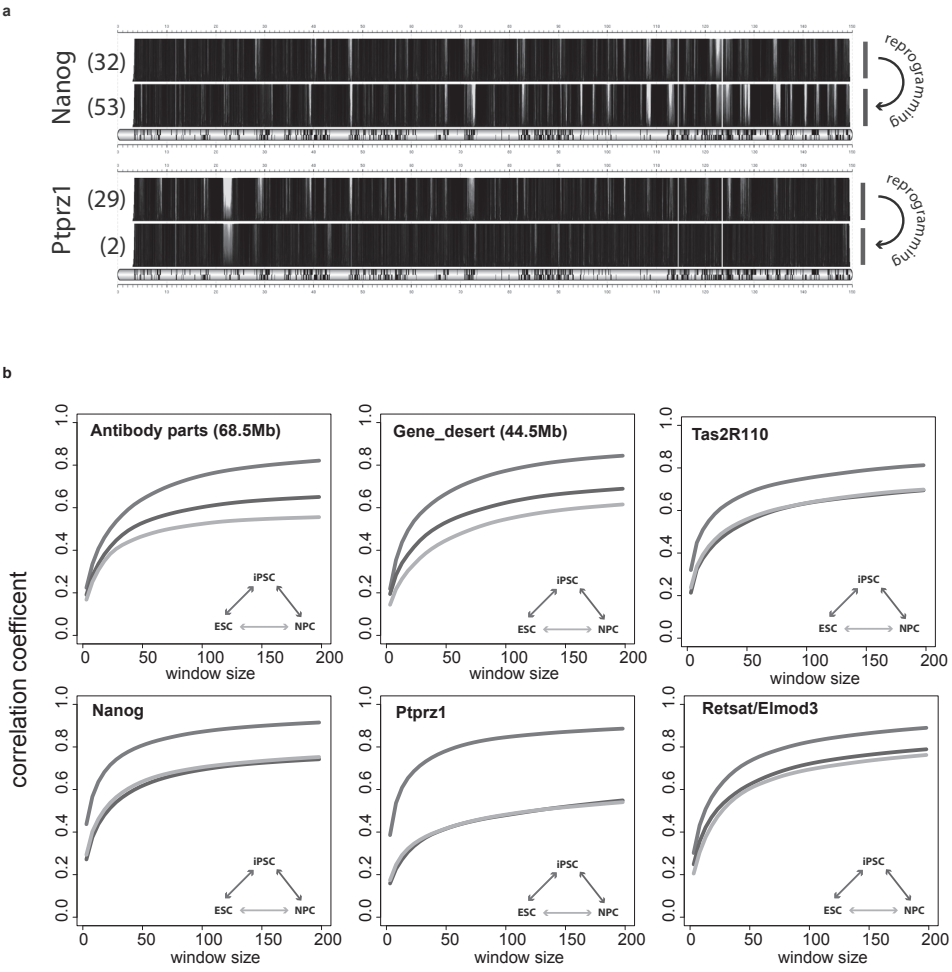
Supplementary Figure 7. Virtual 4C plots show strong similarity between 4C and HiC data. A) Virtual 4C plots were generated by averaging normalized HiC data over windows of 100 kb. Virtual 4C data is compared to actual 4C data also averaged over windows of 100 kb. B) Smoothed scatterplots and Spearman correlation show the high degree of similarity between the 4C and the HiC data.



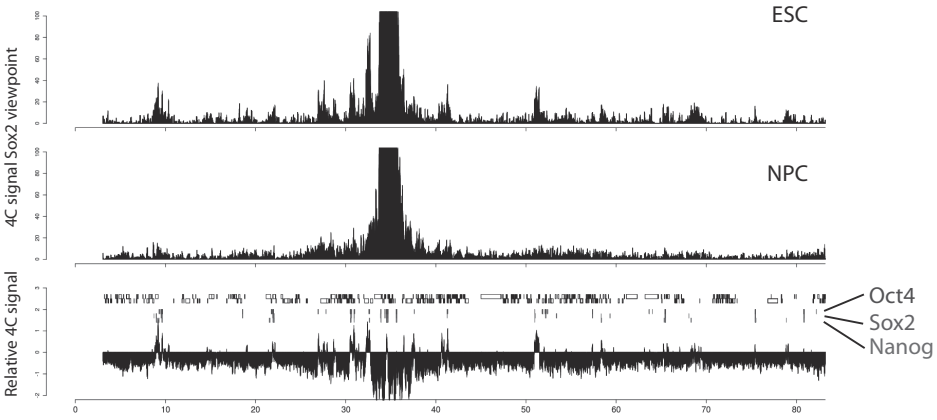
Supplementary Figure 8. Genome-wide quantification of disorganization. Disorganization scores are calculated in a similar way as in Figure 1g.



Supplementary Figure 9. Expression analysis of NPC markers and pluripotency markers. A) qPCR analysis of Nanog mRNA levels using intron spanning primers. B) qPCR analysis shows upregulation of ESC marker genes in iPSCs to ESC levels. Expression values are normalized to NPC levels. C) qPCR analysis as in A) but with marker genes for neural lineage.

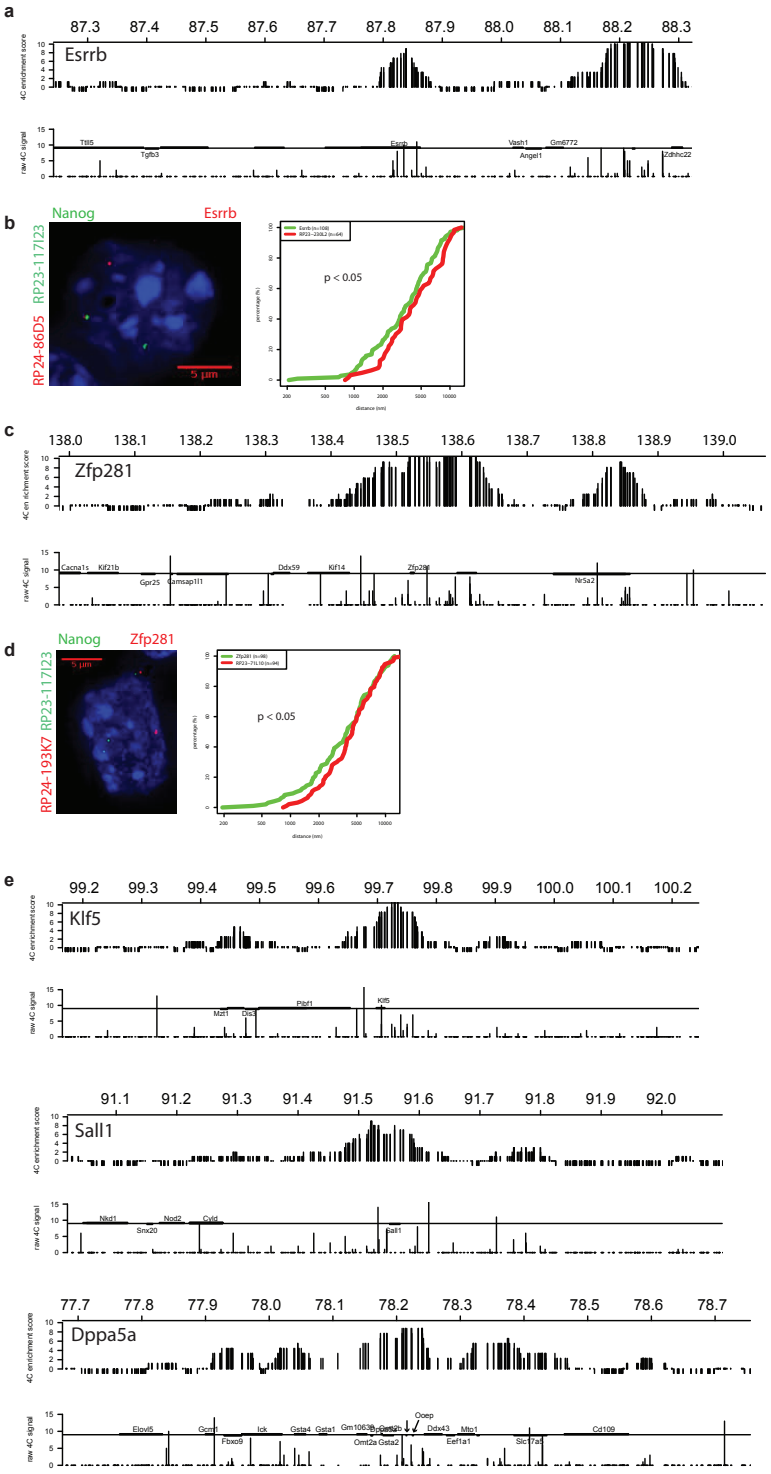


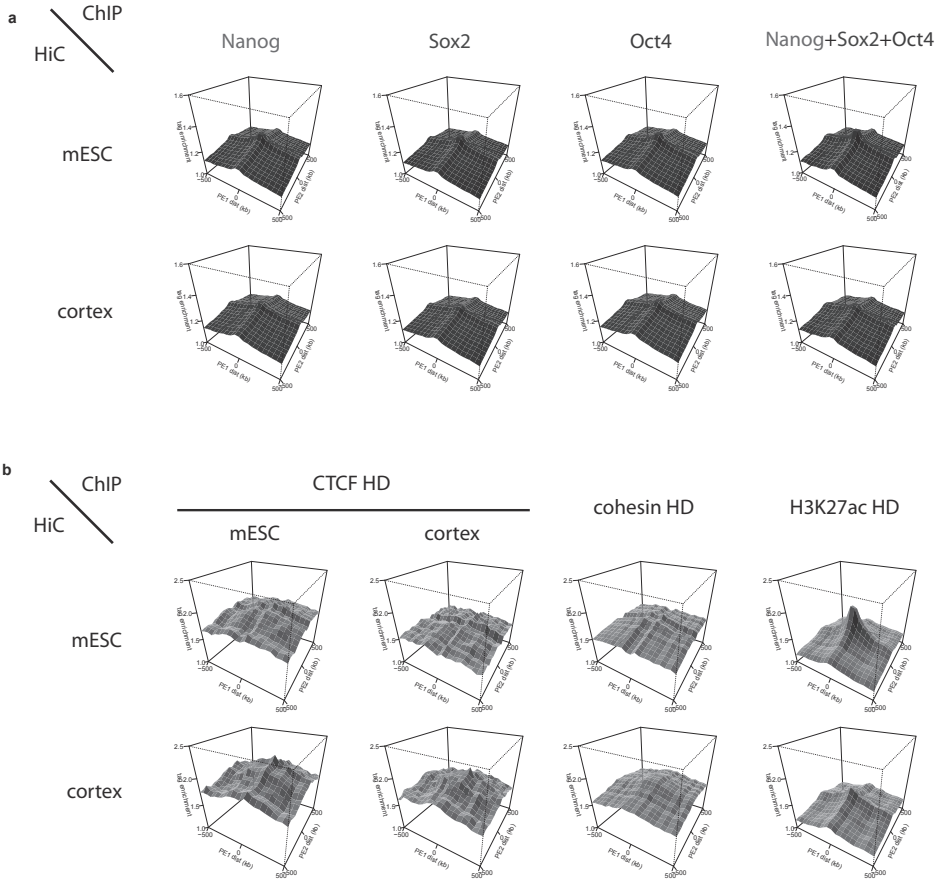
Supplementary Figure 10. Reprogrammed cells have an ES cell like topology. A) Domainograms show 4C profiles of the *Nanog* and *Ptprz1* viewpoint in NPCs and iPSCs; 4C signal as in Figure 1a. B) Windowed Spearman correlation analysis for panel of chromosome 6 viewpoints. Shows the pairwise similarity in 4C profiles between ES, iPS and NP cells. 4C signal is calculated as in Figure 1A with window sizes ranging between 5 and 200. Colored lines show the correlations of iPSCs with ESCs (red) or NP cells (purple), gray line shows the correlations between ESCs and NPCs.



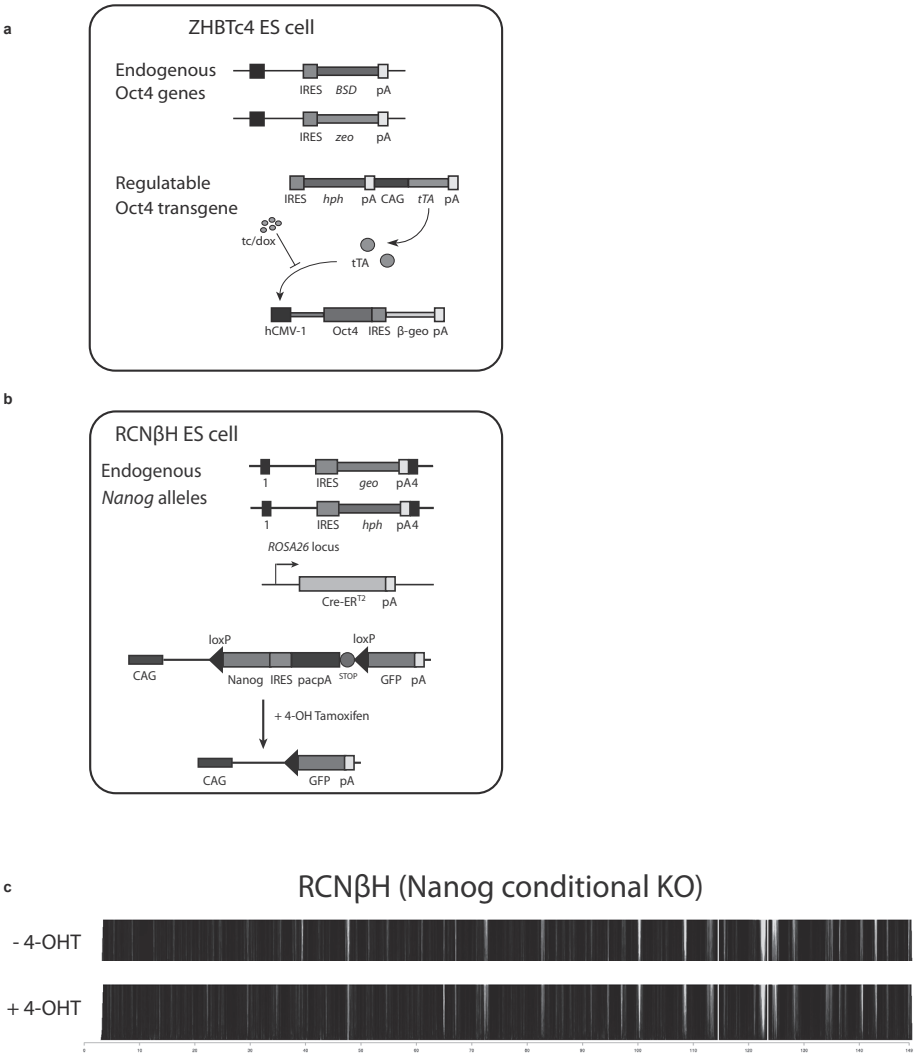
Supplementary Figure 11. *Sox2* contacted regions are enriched for Nanog/Oct4/Sox2 binding sites. Sox2 contact profiles are plotted in the same manner as in Figure 2a.

Supplementary Figure 12. Interchromosomal contacts involving pluripotency genes. A, C) For *Esrrb* and *Zfp281* the 4C enrichment score (see Methods) is shown. Below the windowed data the raw sequencing counts are plotted. B,D) Two-color 3D-DNA-FISH was performed to verify the interchromosomal contacts shown in A) and C). A BAC overlapping with *Esrrb* (RP24-86D5) was compared to a BAC overlapping with the *IgH* locus (RP23-230L2, which is not contacted by *Nanog*). A BAC overlapping with *Zfp281* (RP24-193K7) was compared to a BAC overlapping with a flanking gene desert (RP23-71L10). In all cases RP23-117I23 was used as the BAC that identifies the *Nanog* gene. Monte Carlo testing for spots within 1.5 µm revealed a significant difference between the contacted (green) and the non-contacted (red) genomic regions. E) Additional examples of interchromosomal contacts with pluripotency genes.

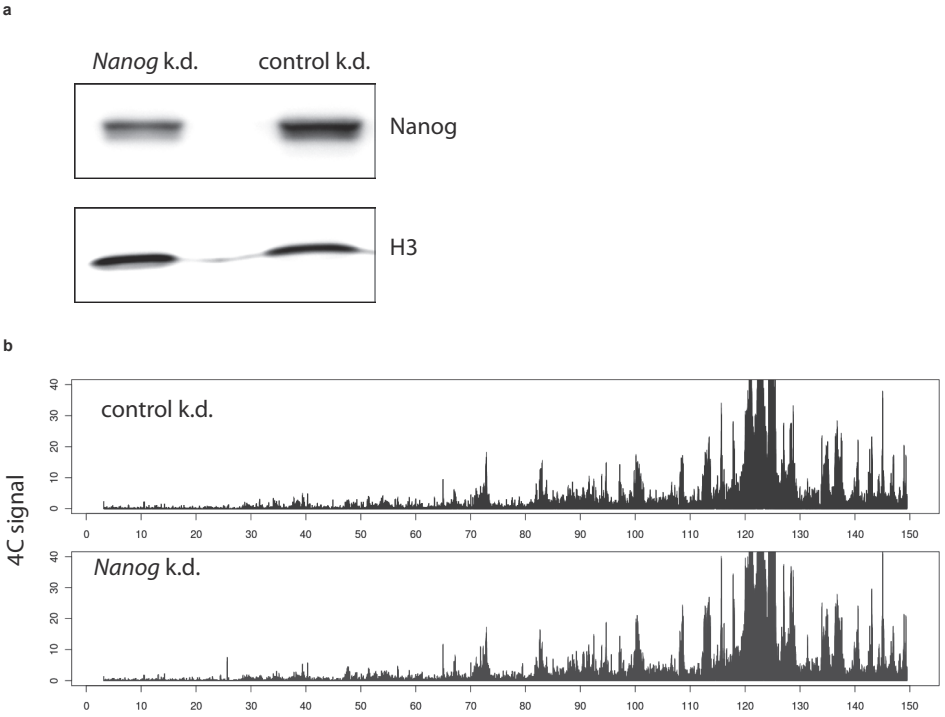

a
Esrrb
b
Nanog
Esrrb
RP24-86D5
RP23-117I23
5 µm
p < 0.05
c
Zfp281
d
Nanog
Zfp281
RP24-193K7
RP23-117I23
5 µm
p < 0.05
e
Klf5
Sall1
Dppa5a



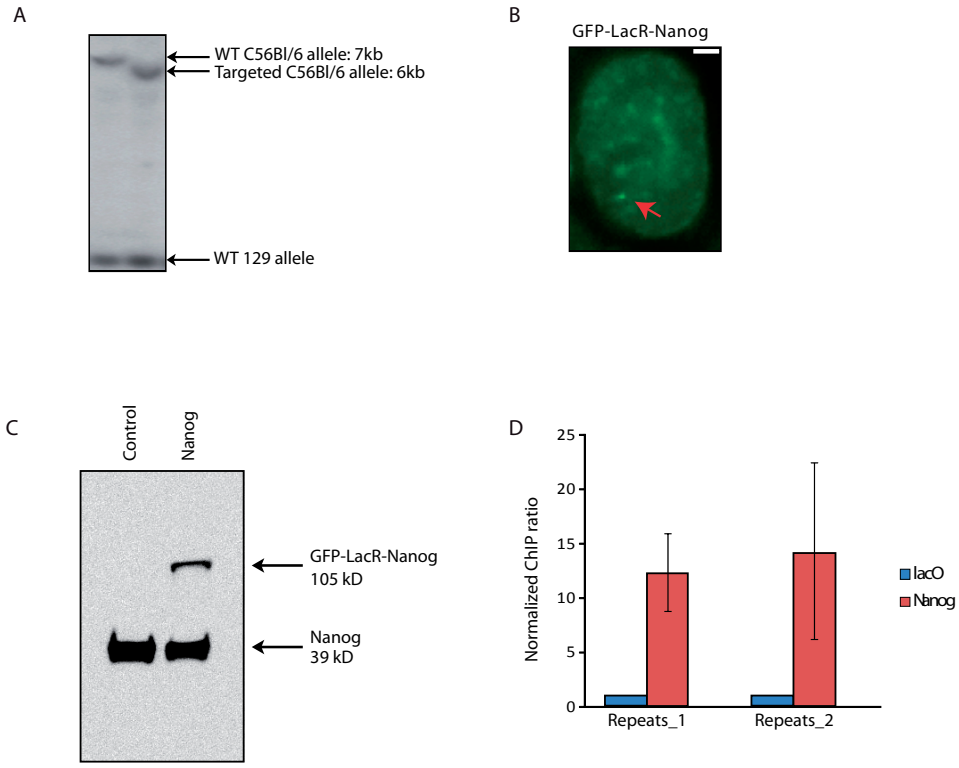
Supplementary Figure 13. PE-SCAn plots for pluripotency factors and high-density clusters of looping factors and enhancers. A) PE-SCAn was performed for single binding sites of Nanog, Sox2, Oct4 and sites where all three bind together. B) PE-SCAn plots are shown for high-density clusters of CTCF (in mESC and cortex), cohesin (Smc1) and active enhancer marks (H3K27ac).



Supplementary Figure 14. Conditional deletion of Nanog does not lead to large differences in overall chromosome topology. A) Schematic representation of ZHBTc4 ES cells with regulatable Oct4 depletion (figure adapted from [17]). Endogenous *Oct4* alleles have been disrupted by targeted substitution with b-geo and hph of sequences extending from within intron I to the 3'UTR. Cells contain a tetracycline/doxycycline-regulated Oct4 transgene that is active in the absence of tetracycline/doxycycline. B) Schematic representation of tamoxifen-treated RCNβH (Nanog$^{-/-}$) ES cells (figure adapted from [18]). Endogenous *Nanog* alleles have been disrupted by targeted replacement of both coding sequences with *β-geo* and *hph*. Cells carry a Cre-ER^{T2} expression cassette inserted into the *Rosa26* locus, and a constitutively expressed cre-revertible *Nanog* transgene under control of the CAG promoter. Cre activation by tamoxifen treatment leads to excision of *Nanog*, which brings GFP under control of CAG and therefore creates green cells as an indication of emergence of Nanog$^{-/-}$ cells. C) Domainograms show the contact profile for RCNβH cells treated with and without tamoxifen, which are depleted for Nanog or expressing Nanog, respectively.



Supplementary Figure 15. Knockdown of Nanog does not lead to drastically altered contact map of the *Nanog* gene. A) Western blot shows the reduction in protein level (78%) in the *Nanog* siRNA knockdown experiment compared to the control knockdown experiment. B) 4C profile for the *Nanog* gene in the control experiment (blue) and the *Nanog* knockdown experiment.



Supplementary Figure 16. Creation and verification of lacO targeted cell lines. A) Southern blot with a probe outside the homology arms (BglII digestion) shows site-specific targeting of 256x lacO repeats into a gene-dense region of chromosome 8. Targeted C56Bl/6 allele becomes 1kb smaller than WT C56Bl/6 allele. Note the 129 allele is present in untargeted and targeted clones. B) Still of a live-cell imaging experiment of lacO ESCs tranduced with a GFP-LacR-Nanog construct. Scale bar indicates 2 μM. Arrow shows the lacO locus bound by GFP-LacR-Nanog. C) Western blot showing endogenous Nanog and GFP-LacR-Nanog by using anti-Nanog antibody in GFP-LacR-Nanog transduced cells. D) ChIP with anti-GFP antibody on ES lacO cells and GFP-LacR-Nanog transduced lacO ES cells. Data was normalized against input and expressed as enrichment over *Hprt*.

Local compartment changes and regulatory landscape alterations in histone H1 depleted cells

Geert Geeven[1#], Yun Zhu[1#], Byung Ju Kim[2#], Boris A. Bartholdy[2#], Seung-Min Yang[2], Todd S. Macfarlan[3], Wesley D. Gifford[4], Samuel L. Pfaff[4], Marjon J. A. M. Verstegen[1], Patrick J. Wijchers[1], John A. Stamatoyannopoulos[5], Arthur I. Skoultchi[2*], & Wouter de Laat[1*]

# equal contribution

[1] Hubrecht Institute-KNAW & University Medical Center Utrecht, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands.

[2] Department of Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA.

[3] Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, Maryland 20892, USA.

[4] Gene Expression Laboratory and the Howard Hughes Medical Institute, The Salk Institute for Biological Studies, 10010 North Torrey Pines, La Jolla, CA, 92037, USA

[5] Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA [2] Department of Medicine, Division of Oncology, University of Washington, Seattle, Washington 98195, USA.

* *UNDER REVIEW*

ABSTRACT

Background

Linker histone H1 is a core chromatin component that binds to nucleosome core particles and the linker DNA between nucleosomes. It has been implicated in chromatin compaction and gene regulation and is anticipated to play a role in higher-order genome structure. Here we have used a combination of genome-wide approaches including DNA methylation, histone modification and DNaseI hypersensitivity profiling as well as Hi-C to investigate the impact of reduced cellular levels of histone H1 in embryonic stem cells on chromatin folding and function.

Results

We find that depletion of histone H1 changes the epigenetic signature of thousands of potential regulatory sites across the genome. Many of them show cooperative loss or gain of multiple chromatin marks. Epigenetic alterations cluster to gene-dense topologically associated domains (TADs) that already showed a high density of corresponding chromatin features. Genome organization at the three-dimensional level is largely intact, but we find changes in the structural segmentation of chromosomes specifically for the epigenetically most modified TADs.

Conclusions

Our data shows that cells require normal histone H1 levels to expose their proper regulatory landscape. Reducing the levels of histone H1 results in a remarkably restricted pattern of massive epigenetic changes and altered topological organization specifically at the most active chromosomal domains. Changes in TAD configuration coincide with epigenetic landscape changes but not with transcriptional output changes, supporting the emerging concept that transcriptional control and nuclear positioning of TADs are not causally related but independently controlled by the locally associated trans-acting factors.

Keywords

Histone H1, chromatin, chromosome conformation, Hi-C, epigenomics

BACKGROUND

DNA in the eukaryotic nucleus is packaged into arrays of nucleosome core particles that are the basic unit of chromatin [1, 2]. Each nucleosome consists of an octamer of four core histones (H2A, H2B, H3 and H4) around which about 145bp of DNA is wrapped. Chromatin also contains a fifth histone, the linker histone, usually referred to as H1. H1 binds to nucleosome core particles near the DNA entry/exit position and to the linker DNA between core particles, stabilizing the association of core particles and DNA and facilitating folding of oligonucleosome arrays into compact structures. Mice and humans express 11 H1 subtypes, including H1a to H1e found at varying levels in most cell types, a replacement subtype (H1(0)) generally associated with terminal differentiation and quiescent states, four germ cell - specific H1's (H1t, H1T2, H1LS1 and H1oo) and a less well-studied subtype (H1x) [3-5]. In addition to their different developmental expression patterns and abundance, the amino acid sequences of these H1 subtypes differ significantly. Despite these differences, we [6-8] and others [9-11] found that elimination of any one of several H1 subtypes and even some pairs of subtypes, does not noticeably affect mouse development. The absence of phenotypes in these mice appears to be due to up-regulation of the remaining subtypes, resulting in maintenance of a normal H1 to nucleosome core particle stoichiometry. However, elimination of three H1 subtypes (H1c, H1d, and H1e) together led to a 50% reduction in the ratio of H1 to core particles and embryonic lethality. Embryonic stem (ES) cells derived from the H1c, H1d, H1e null embryos are viable and also exhibit a 50% reduction in H1: core histone stoichiometry. They show a decrease in the average spacing between nucleosome core particles of about 15 bp, from ~189 bp in normal ES cells to ~174 bp in the triple H1 knock-out (TKO) ES cells [12]. These TKO ES cells also showed decreased local chromatin compaction and selective changes in gene expression. Importantly, up-regulation of certain imprinted and X chromosome-linked genes were prominent and found to be due to H1-dependent alterations in DNA methylation and H3 histone methylation at the regulatory regions of the affected gene [12-14]. Reduced H1 levels were also found to enable CTCF to bind to normally occluded DNA sequences at some imprinted gene loci [13]. CTCF is central factor in setting up local chromatin loops and defining structural domains across mammalian chromosomes [15, 16]. Collectively, this suggests that histone H1 may also have an important function in shaping higher order genome structures in vivo, either directly through its capacity to compact DNA or indirectly by controlling the DNA accessibility of chromatin architectural proteins.

To investigate the role of H1 in genome-wide, higher order chromatin structure, we studied the regulatory landscape and overall genome conformation in the H1-depleted TKO ES cells [12]. Embryonic stem cells show several unique features of nuclear organization as compared to somatic cells. For example, ES cells display hypermobility of chromatin proteins including the core histones and histone H1, indicative of their loosened binding to DNA [17]. Restricting the dynamic state of these core chromatin components compromises the differentiation capacity of ES cells, suggesting that this feature is essential for ES cell identity [17, 18]. ES cells also have an unusually low H1 to nucleosome core stoichiometry:

whereas this ratio is typically 0.75 or more in differentiated cells, in wild-type ES cells it is only about 0.5 [19]. In the H1 TKO ES cells, this ratio is further reduced to 1 histone H1 molecule per 4 nucleosomes [20]. ES cells also display a distinctively disorganized 3D genome with particularly inactive chromosomal regions that fail to cluster as efficiently as seen in somatic cells [21]. Here we applied Hi-C [22] and other genome-wide approaches for mapping epigenetic features to compare wild-type and H1-depleted ES cells to better understand how the H1 linker histone impacts the regulatory and three-dimensional landscape of the genome and its consequences on transcription.

RESULTS AND DISCUSSION

Clustered DNA methylation changes in histone H1 depleted ES cells

Local DNA demethylation was previously observed in the H1-depleted TKO ES cells particularly at the imprinting control regions (ICRs) of H19-Igf2 and Gtl2-Dlk1 loci [12] and at the Rhox 5 promoter on the X chromosome [14]. The activity of H1 in promoting DNA methylation at the ICRs was later attributed to H1's function in recruiting the DNA methyltransferases DNMT1 and DNMT3B [13]. No global changes in DNA methylation were observed at that time, based on digestion with methylation-sensitive restriction enzymes [12]. To study genome-wide methylation changes in more detail, here we used the HELP-tagging assay that enables high-throughput identification of sequences neighboring restriction sites of the methylation-sensitive HpaII enzyme [23]. Approximately 15,000 sites displayed differential methylation across the genome. Consistent with a function of H1 in recruiting DNA methyltransferases [13], more than two-thirds of these sites appeared hypo-methylated in TKO cells, while ~30% were hypermethylated. The differentially methylated sites were not uniformly distributed across the genome but appeared significantly clustered (Figure 1A-B), with the sex chromosomes being surprisingly protected against methylation changes (Additional file 1: Figure S1). To further delineate the clustered distribution, we investigated the density of differentially methylated sites in topologically associated domains (TADs). We considered TADs as chromosomal units of interest as they mark genomic segments within which sequences preferentially contact each other. As such TADs are believed to be the structural and functional genomic units that encompass the genes and their cognate regulatory sites [15, 24]. We categorized TADs according to gene content and created five bins with equal numbers of TADs; each bin had a similar average TAD size, but the gene-poorest bin encompassed only 3-4% of all genes while the gene-richest bin carried over 50% (Figure 1C). Overall, DNA methylation density followed gene density (Figure 1C). Taking this distribution into account we found that changes in DNA methylation significantly clustered in the most gene-rich TADs, no matter whether we considered hyper- or hypo-methylation events (Figure 1D). Collectively, the data shows that DNA methylation changes preferentially occur in chromosomal segments with a high gene density.

To investigate whether the DNA methylation changes associated with H1 depletion localized to any specific type of chromatin we used published ESC ChIP-seq profiles and

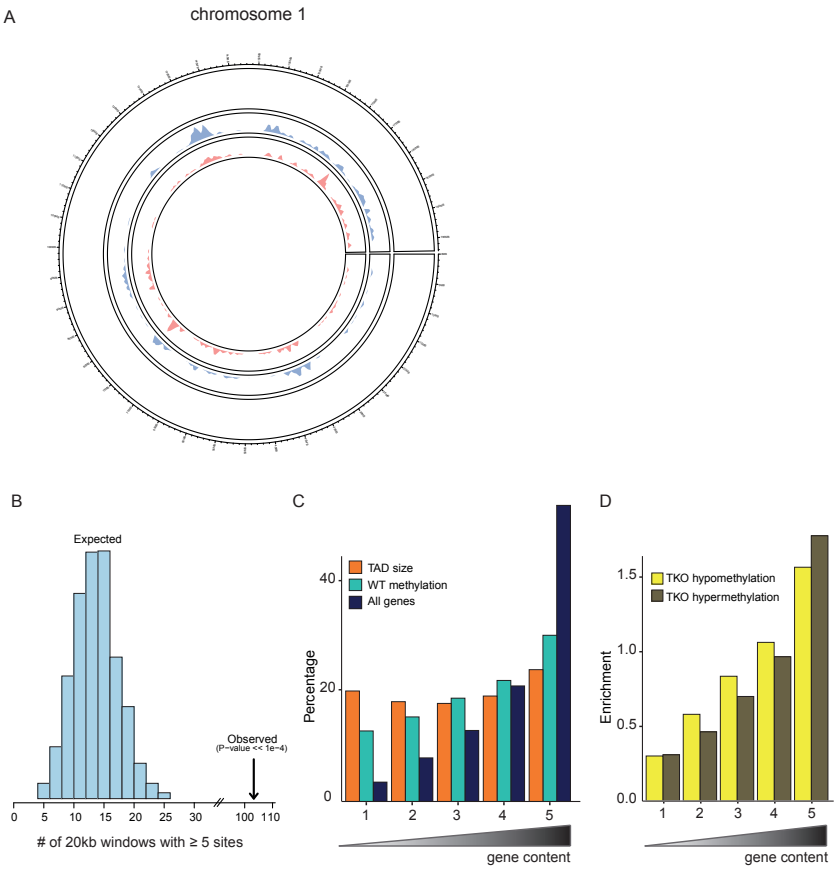


Figure 1. Clustered DNA methylation changes in histone H1 depleted ES cells.

A. Circos plot showing the genome-wide distribution of hypomethylated (in blue) and hypermethylated (in red) loci along the linear sequence of chromosome 1 in H1 TKO compared to wildtype ES cells, based on the HELP-tagging assay.

B. Histogram of counts of number of genomic windows (fixed size, 20Kb, non-overlapping) that contain at least 5 randomly chosen sites out of the complete set of roughly one million assayable sites in the HELP-tagging assay. The counts in the histogram sum to the total number of random draws, i.e. 1000. The arrow indicates the observed number (103) of genomic windows that contain at least 5 differentially methylated sites, which is significantly more than expected by chance.

C. Barplots showing the percentages of genes compared to the percentages of DNA methylated sites in wildtype ES cells in groups of TADs ranked according to the number of overlapping genes (gene content). The ranking on the x-axis is such that the leftmost bin contains the 20% TADs with the lowest number of genes and the rightmost bin TADs with the highest number of genes. The genomic sizes of the groups of TADs as a percentage of the total genomic size of all TADs is plotted as a reference.

D. Barplots showing the ratios of both the percentage of hypermethylated or hypomethylated sites in TKO cells over the percentage of DNA methylated sites in WT ES cells in the sane groups of TADs as defined before (Figure 1D). The ratio thus measures the amount of enrichment or depletion of hyper- and hypomethylation in TKO cells in each bin.

a partitioning of the genome into different chromatin states by means of ChromHMM, a segmentation algorithm for the identification of chromatin states based on the presence of combinations of chromatin modifications [25]. We found that particularly predicted enhancers were strongly overrepresented among hypomethylated CpGs (Additional file 2: Figure S2), which further indicates that histone H1 plays a role in controlling the DNA methylation status at potential regulatory sequences.

Altered genomic regulatory landscape in H1 TKO cells

To further characterize the consequences of histone H1 depletion on the regulatory chromatin landscape of ESCs we determined the distribution of DNase hypersensitivity (DHS), H3K4me1 and H3K4me3 sites across the genome of H1 TKO ESC lines and their wildtype counterparts. Roughly identical numbers of DHS were scored in the two conditions (281934 sites in TKO vs 293319 sites in WT, medians over triplicates), and genome-wide DHS patterns between WT and TKO were very similar both to each other and to DHS profiles determined in another ES cell line by the ENCODE project (Additional file 3: Figure S3). When ranked according to their delta DHS signal, *de novo* formed DHS were clearly appreciable, but there was little evidence for complete loss of DHS in H1 TKO cells (Figure 2A). Chromatin-associated histone H1 was previously found to interfere with the binding of histone methyltransferase SET7/9, thereby preventing methylation of H3K4 [13]. ChIP-seq results revealed no change in the overall numbers of H3K4me1 sites across the genome but did show a high number of sites that gained (6536) or lost (7319) mono-methylation, indicating dynamic changes in this enhancer mark (Additional file 4: Figure S4A). When looking at the H3K4me3 ChIP-seq results we found four times more sites with increased than with decreased levels of trimethyation (2043 versus 495) (Additional file 4: Figure S4B).

We next wished to understand the relationship between these epigenetic changes. Since differences in DHS were clearest for the 2123 newly formed DHS, we focused on those DHS and asked whether their formation coincided with other epigenetic changes. Nearly one third of the new DHS sites also showed a gain in either H3K4me1 or H3K4me3 or both, whereas loss of these marks was very infrequently observed at new DHS sites (Figure 2B). More than 10% (256 / 2123) of the new DHS sites also revealed loss of DNA methylation, while the opposite, hypermethylation, was rarely found at these sites (19 times) (Additional file 5: Figure S5). When focusing on H3K4me3 sites, those with increased H3k4me3 levels often (>25%) also showed a gain (and seldom a loss) in H3K4me1, while sites losing H3K4me3 frequently showed concomitant loss of H3K4me1 (also >25%) (Figure 2C). Finally, when considering the differential methylated CpGs, sites with reduced methylation in TKO were often enriched for H3K4me1 marks in these cells and, vice versa, hypermethylated sites frequently lost H3K4me1 (Figure 2D).

In summary, the depletion of histone H1 changes the chromatin signature of thousands of sites across the genome. Many of them show concomitant loss or gain of multiple chromatin marks. Nearly invariably these combinatorial changes are either all activating or all repressing, implying that they can cooperate to strengthen or dampen the regulatory potential of a site.

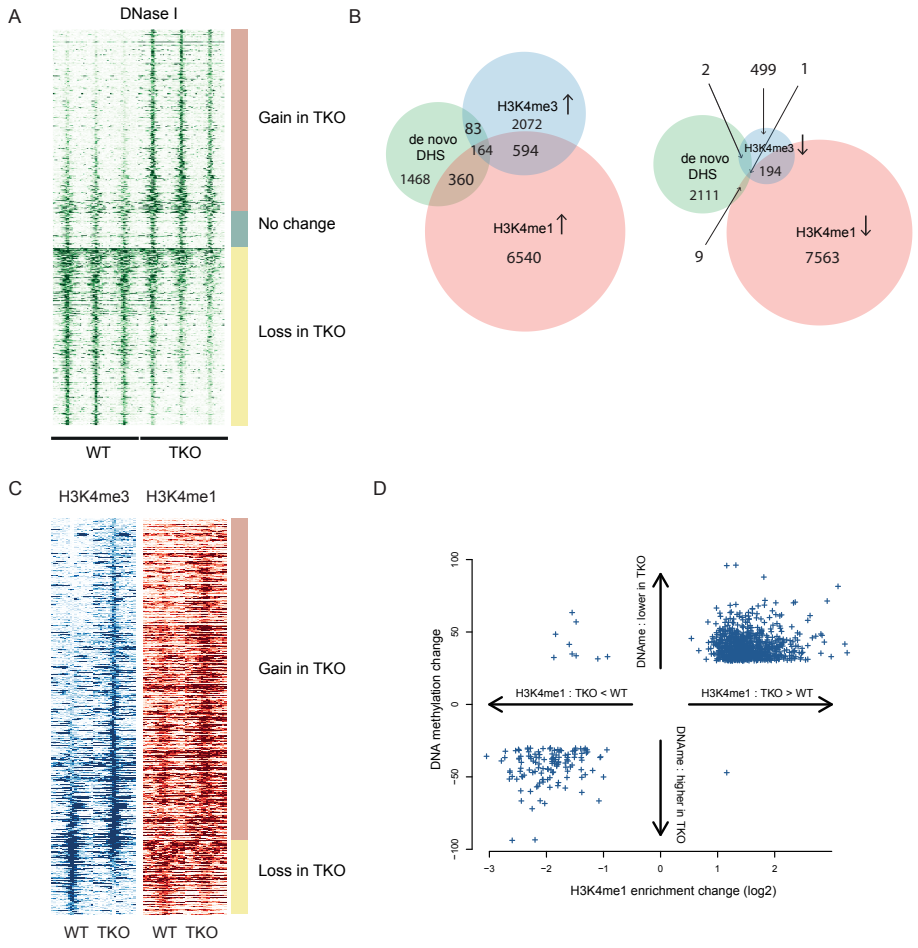


Figure 2. Altered genomic regulatory landscape in H1 TKO cells

A. Heatmap of DNase-seq coverage in triplicate experiments in wildtype and H1 TKO cells. Genome-wide statistical analysis of differences in DNase-seq coverage between wildtype and TKO revealed 2123 sites (top) with a gain and 2043 sites (bottom) with a loss of DNase I hypersensitivity in TKO cells respectively. The rows of the heatmap correspond to these ~4000 sites, ordered by log-fold change in coverage, with a random collection of unaltered DNase I hypersensitive sites in between.

B. Venn diagrams that contain counts of sites that gain (left) or loose (right) enrichment of the histone marks H3K4me1 and H3K4me3 and their overlap with the 2123 newly formed DHS sites in TKO cells.

C. Heatmaps of ChIP-seq enrichment for histone marks H3K4me1 and H3K4me3 in wildtype and H1 TKO cells. Profiles represent averages over duplicate experiments. Genomic sites represented by rows in the heatmap are sites where significant changes in H3K4me3 enrichment are observed. Rows are ranked by the magnitude of that change from top to bottom in descending order of increase in H3K4me3 enrichment in TKO cells.

D. Scatter plot of changes in H3K4me1 enrichment and DNA methylation at sites where significant changes in both are observed.

Epigenetic changes accumulate in gene-dense TADs

We then wished to understand where in the genome these epigenetic changes take place. For this we again considered TADs as the genomic units of interest and we intersected the various datasets with the previously defined five classes of TADs. Not surprisingly, the general distribution of DHS, H3K4me1 and H3K4me3 sites in wildtype (and TKO) cells closely followed that of genes, with all of these marks specifically accumulating in the most gene-dense TADs (Additional file 6: Figure S6). When correcting for their overall distribution there was no obvious enrichment of sites losing DHS or H3K4me3 signal in any of the TAD bins (Figure 3A). This suggests that sites showing loss of hypersensitivity or loss of the promoter-mark H3K4me3 are distributed proportional to the overall genomic localization of DHS and H3K4me3 sites. In contrast, sites losing H3K4 monomethylation in H1 TKO cells were significantly depleted (chi-squared test $P << 10^{-6}$) from the most gene-dense TADs and appeared accumulating in the gene-poorest TADs (Figure 3B). This could indicate that normal histone H1 levels are needed for proper maintenance of H3K4me1 levels in inactive chromatin surroundings. Alternatively, H3K4me1 sites in active chromatin surroundings are relatively protected against demethylation. To further investigate if epigenetic alterations occurred at specific genomic locations we looked at de novo acquired active chromatin marks. We defined de novo DHS sites as those that were exclusively identified in TKO cells but also lacked threshold H3K4me1 or H3K4me3 levels in wildtype cells. Similarly, we defined de novo formed H3K4me1 sites as those scoring positive for this mark only in TKO cells and also lacking significant H3K4me3 and DHS signal in wildtype cells. These non-bookmarked sites are ubiquitously present and their conversion into active sites may therefore a priori take place anywhere in the genome. However, new hypersensitive sites and new H3K4me1 sites both preferentially accumulated again in the most gene dense TADs (Figure 3C-D). Thus, despite being a generic chromatin component present throughout the genome, depletion of H1 results in a preferential gain of active chromatin marks within the most gene-dense TADs. These TADs are already dense in such regulatory chromatin signatures and this, we speculate, may create sensitized chromatin that is exquisitely susceptible to further epigenetic changes. An alternative, not mutually exclusive, explanation is that these TADs form nuclear compartments where the corresponding chromatin modifying enzymes accumulate to cooperatively establish and maintain the required dense landscape of regulatory sites. Perturbing the integrity of the chromatin template, as occurs when H1 histone levels are reduced, may further stimulate local mass action and increase the chance to modify neighboring chromatin sites. Precedence for local cooperative action exists: it was recently shown in *Drosophila* that clustered low affinity binding sites better accumulate PcG proteins than their more isolated counterparts elsewhere in the genome [26].

Genes with altered expression are proportionally distributed across the genome

To investigate how the altered regulatory chromatin landscape functionally translates into gene expression changes we assayed the genome-wide transcriptome. RNA-seq confirmed previous observations obtained by microarray analysis. We again found a small subset of genes (<200) that showed a significant 2-fold or higher change in expression, the majority

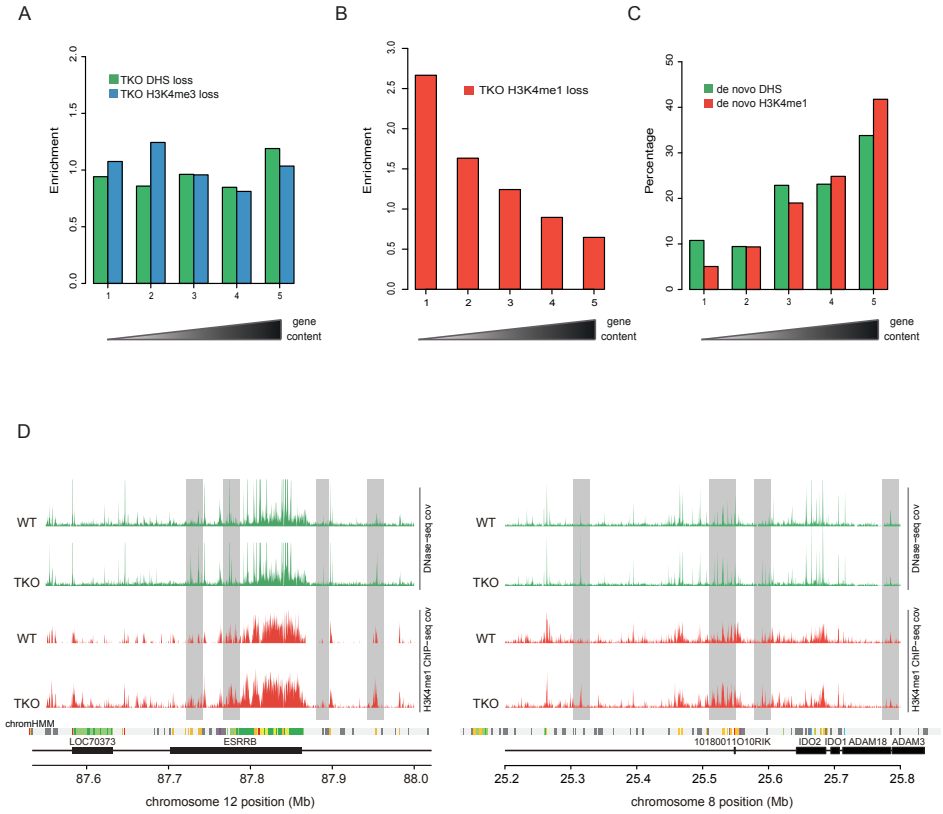


Figure 3. Epigenetic changes accumulate in gene-dense TADs

A. Barplots showing the ratio of (the percentage of) sites with a significant loss of DHS in TKO cells, over the (percentage of) DHS sites in WT ES cells in groups of TADs. The TADs are ordered based on gene content and grouped in equally sized bins (same ordering as in Figure 1D), with the most gene poor TADs on the left. An analogous ratio is plotted for the sites that loose H3K4me3 in TKOs, but here the ratio is computed relative to the WT sites with H3K4me3 enrichment.

B. Same as panel A, but for sites that significantly loose H3K4me1 enrichment in TKO cells (with the ratio compared relative to the WT H3K4me1 sites).

C. Barplots showing the percentages of de novo DHS sites in groups of TADs ranked according to the number of overlapping genes (same ranking as in panel (A,B)). Also shown are the percentages of de novo H3K4me1 sites in TKO ES cells.

D. Example of 2 loci, one on chr12 and one on chr8, where several novel DHSs appear which co-occur with changes in H3K4me1 in TKO ES cells. Normalized DNase-seq coverage is plotted in green (averaged over triplicate experiments in WT and TKO) and normalized H3K4me1 ChIP-seq coverage is plotted in red (averaged over duplicates). Black boxes indicate genes and a track containing the different computationally predicted chromatin states in WT mouse ES cells (chromHMM) is shown as a reference.

(>75%) of which showed reduced levels of transcription (Figure 4A). Among those were the previously described *Hox* genes [27], while the most prominently upregulated genes included a series of paternally imprinted genes [12] (Figure 4B). Previous detailed characterization of two of the most strongly upregulated loci in TKO cells, the paternally imprinted *Gtl2* locus and the *H19* locus, revealed hypomethylation of their imprinting control regions [13]. To investigate whether changes in DNA methylation generally underlie transcriptome changes we compared the genomic distribution of differentially expressed genes and differentially methylated sites. We again investigated this at the level of TADs. To maximally exploit the benefit of an integrative analysis we considered a less stringent set of 598 differentially expressed genes. We ranked TADs based on the number of overlapping differentially DNA methylated sites and computed the fraction of differentially regulated genes. Figure 4C shows that indeed TADs with most changes in DNA methylation co-segregated with those most enriched for differentially expressed genes. However, the non-uniform genomic distribution of genes appeared a confounding factor here as a similar relationship was found between DMR distribution and gene density in general (Figure 4C). To investigate this in more detail we ranked TADs according to gene content. Indeed this categorization highly correlated with the distribution of differentially expressed genes (Figure 4D), implying that, from a genomic distribution point of view, they are a proportional and apparently random collection of genes.

Thus, while new epigenetic features, de novo acquired in H1 TKO cells, preferentially appeared in TADs that in wildtype cells already carried the highest density of these marks, these same TADs, when normalized for gene content, were not enriched for deregulated genes. It therefore seems that gene regulation is not controlled by the general distribution of regulatory chromatin signatures across TADs. More likely, it depends on the regulatory status of a limited collection of relevant sequence modules that together, possibly via chromatin looping, determine the transcriptional output of a cognate gene.

Higher-order topological changes follow epigenetic but not transcriptional changes

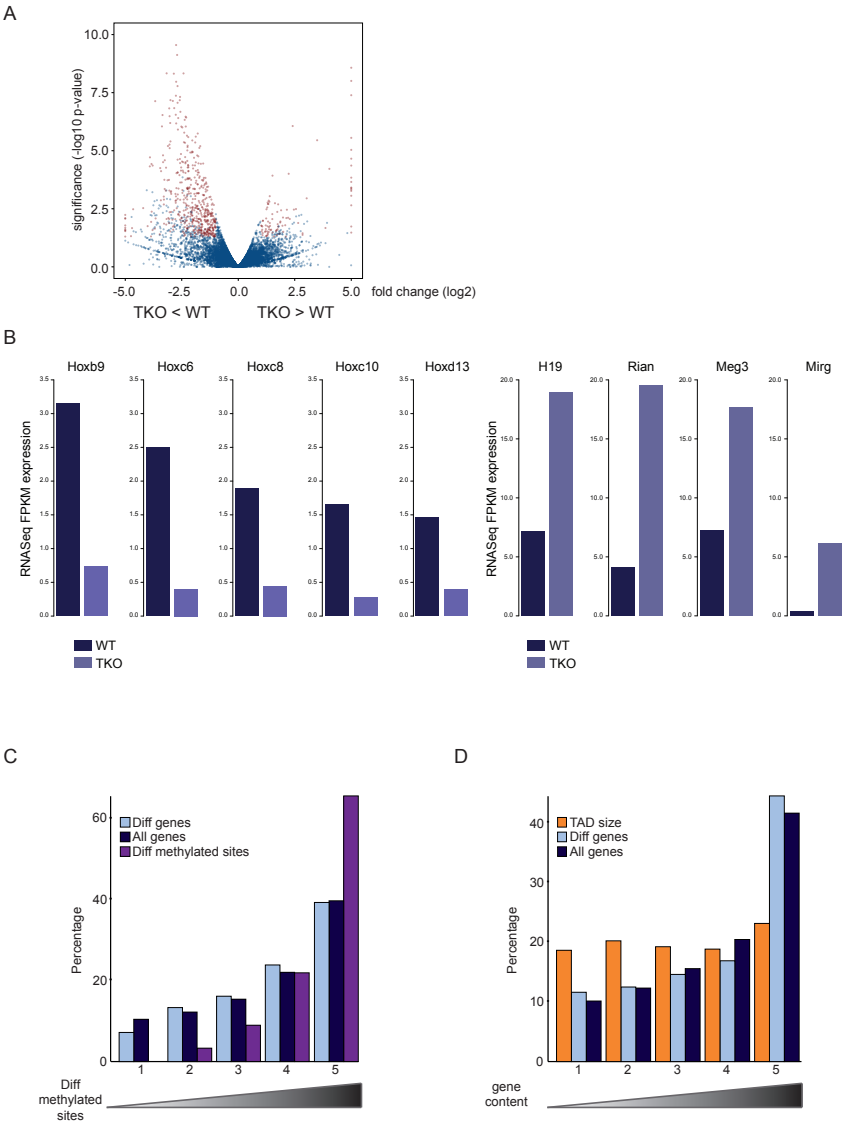
We finally wished to understand how histone H1 depletion and the accompanying epigenome and transcriptome changes affect the overall 3D organization of the genome. To this end we

Figure 4. Genes with altered expression are proportionally distributed across the genome

A. 'Volcano plot' of statistical significance (-log10 p-value) against fold-change comparing RNA-seq gene expression between WT and H1 TKO mouse ES cells. Transcripts that are significantly differentially expressed between the two conditions are shown in red, whereas the genes in blue do not reach the threshold.

B. Barplot showing RNA-seq normalized expression values for a selection of transcripts previously reported to be down-regulated (*Hox* genes, left panel) and up-regulated (imprinted genes, right panel) in H1 depleted ES cells.

C. Barplots showing the percentages of differentially expressed genes compared to all genes and compared to the percentages of differentially DNA methylated sites in TKO ES cells. Percentages are calculated in groups of TADs ranked according to the number of overlapping differentially ▶



methylated sites. The ranking on the x-axis is such that the leftmost group contains the 20% TADs with the lowest number of differentially methylated sites and the rightmost group of TADs contain the highest number of such sites.

D. Barplots showing the percentages of differentially expressed genes compared to all genes in groups of TADs ranked according to the total sum of all epigenetic changes. This sum is calculated by summing the changes in DNA methylation, the histone marks H3K4me1 and H3K4me3 and the number of differential DHS sites in each TAD. The ranking on the x-axis is such that the leftmost group contains the 20% TADs with the lowest number of changes and the rightmost group of TADs contain the highest number. The genomic sizes of the groups of TADs as a percentage of the total genomic size of all TADs is plotted as a reference.

performed replicate Hi-C experiments, each with a different, frequently cutting, restriction enzyme (NlaIII and DpnII) [28], in both TKO and matched wildtype ES cells. Each dataset contained between 26M-42M valid Hi-C read pairs, adding up to 53M (wildtype) and 76M (TKO) valid Hi-C read pairs per cell type. All Hi-C libraries showed an equal high ratio of intra- over inter-chromosomal contacts (~75%), indicative of good quality Hi-C libraries [29]. We normalized and processed Hi-C data by binning reads per 100 kb chromosomal segments to generate contact heatmaps as described before [22]. Visual inspection of the heatmaps suggested that chromosomes folded very similar between wildtype and TKO cells (Figure 5A). Principal component analysis of Hi-C data was previously used to uncover an A and B compartment where respectively active and inactive chromatin regions preferentially cluster. When applied to our datasets it showed that chromosomal domain organization and overall 3D genome structure is indeed very similar between the two cell types (Figure 5A). In fact, the contact profiles that we generated for wildtype and histone H1 depleted ESCs

Figure 5. Higher-order topological changes follow epigenetic but not transcriptional changes ▶

A. Normalized Hi-C interaction heatmap showing chromatin compartmentalization (A/B compartments) at 100kb resolution for chromosome 1 in wildtype (left) versus TKO (right). Coefficients of the first principal component (PC1) of the Hi-C interaction heatmap are plotted on top along the linear sequence of chromosome 1, showing no apparent changes in chromatin compartment organization upon H1 depletion in mouse ES cells.

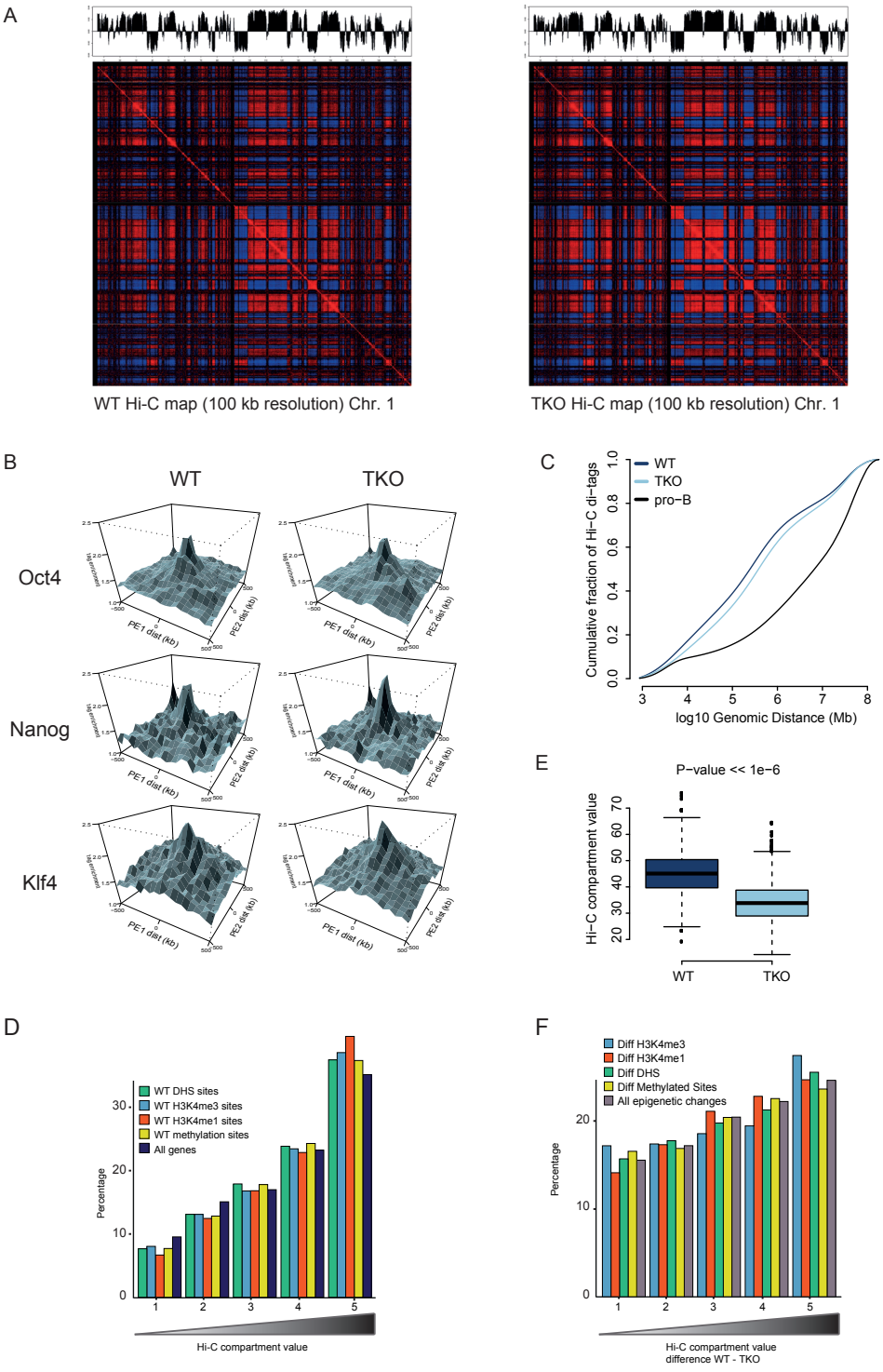
B. PE-SCAN Hi-C analysis probing Hi-C interactions between clusters of binding sites for transcription factors that control mouse ES cell identity (pluripotency factors). ES specific interactions among binding sites of *Oct4*, *Nanog* and *Klf4* remain present in mouse ES cells upon depletion of H1 in triple knock-out cells.

C. Plot comparing the distribution of Hi-C interactions versus genomic distance for three different Hi-C maps. Mouse ES cells are characterized by a relatively large fraction of interactions over short distances, whereas differentiation is known to be accompanied by an increase in long-range interactions. The TKO Hi-C map clearly shows a shift towards those of a more differentiated cell.

D. Barplots showing the percentages of genes in groups of TADs ranked according to the Hi-C compartment value in WT ES cells. The ranking on the x-axis is such that the leftmost group contains the 20% TADs with the lowest and the rightmost group of TADs contains the highest Hi-C compartment values. We also show the distribution of sites, in WT ES cells, enriched for the histone marks H3K4me1and H3K4me3, sites with DNA methylation and DNase I hypersensitive sites in these groups of TADs.

E. Boxplots comparing the Hi-C compartment value of all mouse ESC topological domains in our WT and TKO Hi-C maps. A two-sample Wilcoxon rank sum test was applied to test the significance of the shift in compartment value in TKO cells (P-value << 1e-6).

F. Barplots showing the percentages of the total sum of all epigenetic changes in groups of TADs ranked according to the difference in Hi-C compartment value between TKO and WT ES cells. The ranking on the x-axis is such that the leftmost group contains the 20% TADs with the lowest and the rightmost group of TADs contains the highest difference in Hi-C compartment value. We also show the percentages for the individual changes in the histone marks H3K4me1and H3K4me3, differential DNA methylation and differential DHS.


A
WT Hi-C map (100 kb resolution) Chr. 1
TKO Hi-C map (100 kb resolution) Chr. 1
B
WT
TKO
Oct4
Nanog
Klf4
PE1 dist (kb)
PE2 dist (kb)
C
WT
TKO
pro−B
Cumulative fraction of Hi−C di−tags
log10 Genomic Distance (Mb)
E
P−value << 1e−6
Hi−C compartment value
WT
TKO
D
WT DHS sites
WT H3K4me3 sites
WT H3K4me1 sites
WT methylation sites
All genes
Percentage
Hi-C compartment value
F
Diff H3K4me3
Diff H3K4me1
Diff DHS
Diff Methylated Sites
All epigenetic changes
Percentage
Hi-C compartment value
difference WT - TKO

were more similar to each other than any of the two was to a previously published wildtype ESC contact profile [15] or pro-B cell contact profile [30] (Additional file 7: Figure S7). Thus, a 50% depletion of linker histone H1 can be tolerated without profound changes in the overall 3D genome.

We previously reported that ESCs harbor a unique 3D genome, hallmarked in general by a more random higher order topology with particularly the inactive chromatin compartment being spatially unorganized. Another feature of the pluripotent 3D genome is the specific clustering of genomic regions dense in binding of pluripotency factors [21]. This preferential clustering of dense pluripotency factor associated regions was also appreciable in both our wildtype and H1 TKO cells (Figure 5B) and confirmed that the TKO cells harbored an ES cell identity. Compared to their wildtype counterparts, the TKO cells did show an increased capacity for chromosomal regions to contact each other over distance (Figure 5C). This is reflective of the more strict 3D organization typically seen in differentiated somatic cells, but the effect in H1 TKO cells was not as pronounced as seen in for example differentiated pro-B cells (Figure 5C). Nevertheless, the intermediate phenotype may be indicative of some differentiation taking place in the TKO cells.

Realizing that the overall topology of chromosomes is unaltered in H1 depleted cells but that domains gain capacity to reach out and contact other domains elsewhere on their chromosome, we used the recently published TAD cross-boundary ratio to search for more subtle topological changes. The TAD cross-boundary ratio divides the intra-domain contacts over the inter-domain contacts [31] and as such can serve as a compartment value. It is the most gene-dense TADs that are most enriched for DHS sites and for active chromatin marks which show the highest such compartment value (Figure 5D). Active TADs therefore seem the structurally most isolated chromosomal entities. We calculated this value for each TAD in wildtype and H1 TKO cells and compared them. In agreement with the observation that in histone H1 depleted cells chromosomal sites more easily engage in contacts over very large distances, TADs generally had a lower compartment value in TKO versus wildtype cells (Figure 5E). Histone H1 therefore seems to contribute to the topological segmentation of chromosomes.

We then wished to identify the TADs that upon histone H1 depletion were most sensitive to topological changes. For this we computed the difference in compartment value between wildtype and TKO cells and ranked TADs accordingly. Despite almost all TADs showing a reduction in the compartment value and despite this difference being relatively modest, we nevertheless found that the degree of structural changes significantly correlated (chi-squared test $P << 10^{-6}$) with the amount of epigenetic changes observed per TAD (Figure 5F). This was true no matter which of the investigated marks was considered. Thus, the top 20% topologically most altered TADs were those that also carried most sites with altered hypersensitivity, most sites with modified levels of H3K4me1 and/or H3K4me3 and most differentially methylated CpGs. In contrast, the TADs most resistant to topological changes were those showing the least epigenetic changes upon H1 depletion. Interestingly, such correlations were not found with gene density, TAD size or differential gene expression (Additional file 8: Figure S8). It is therefore not necessarily the TADs with highest gene

content, nor the larger TADs and also not the TADs with most striking changes in transcriptional output that are most sensitive to topological changes. Rather, alterations in the epigenetic landscape appear to best correlate with topological alterations of TADs. Thus, while gene expression clearly correlates with the nuclear positioning of TADs relative to each other and to for example the nuclear periphery, our current data provides further evidence that gene expression and higher-order chromosome topology are not causally related [32-34]. Rather, they may be independently controlled by the locally associated trans-acting factors.

CONCLUSIONS

Our data shows that cells require normal histone H1 levels to expose their proper regulatory landscape. Reducing the levels of histone H1 results in a remarkably restricted pattern of massive epigenetic changes and altered topological organization specifically at the most active chromosomal domains. Changes in TAD configuration coincide with epigenetic landscape changes but not with transcriptional output changes, supporting the emerging concept that transcriptional control and nuclear positioning of TADs are not causally related but independently controlled by the locally associated trans-acting factors.

METHODS

Cell culture

WT and H1 TKO ES cells [12] were grown on irradiated mouse embryonic fibroblasts in DMEM (high glucose, Gibco) with 15% FBS, 1x non-essential amino acids (NEAA; Gibco), 1x penicillin–streptomycin (Gibco), 1:1,000 b-mercaptoethanol (Invitrogen), 1x L-glutamine (Gibco) and 1,000 U/ml leukaemia inhibitory factor (LIF; Gibco).

Hi-C template

Cells were trypsinized and plated on uncoated plates for 30 minutes at 37°C to get rid of the feeder cells. Then 3C template was generated as previous described [35]. In brief, 10 million cells were cross-linked by 2% formaldehyde, then digested with DpnII or NlaIII, and ligated to form 3C circles. Purified 3C products were then further sheared to 600-800 bp. 1ug sheared DNA was used to generate high-throughput sequencing ready sample by using TruSeq DNA sample prep kit (Illumina), following the standard commercial protocol. The Hi-C library was sequenced with Illumina paired end sequencing.

ChIP-seq

Chromatin immunoprecipitation was performed as described previously (Fan Cell 2005; Yang PNAS 2013) with ChIP grade H3K4me1- and H3K4me3-specific antibodies purchased from Abcam with a few modifications. Fixed cells were lysed in a buffer containing 10 mM Na-Butyrate and isolated chromatin was sonicated to 500-800 bp with a Covaris S2 sonicator at 4^0C. ChIP-seq library preparation and sequencing was performed by the Epigenomics Core Facility at the Albert Einstein College of Medicine using on an Illumina

2500 HiSeq instrument. We generated duplicate ChIP-seq libraries (both duplicate input and IP samples, with antibodies against H3K4me1 and H3K4me3) for both conditions (WT and TKO). Reads from all 8 libraries where aligned to the reference genome (NCBI37/mm9) with bowtie [36]. Regions significantly enriched for H3K4me1 and H3K4me3 compared to matched input samples where identified using the MACS peak caller [37] and differential ChIP analysis was done using F-Seq [38].

RNA-seq

Total RNAs were prepared from ES cells adapted to gelatinized dishes using RiboPure RNA purification kits (Ambion). Paired end library construction was performed using Tru-seq kits (Illumina). Resulting libraries were run on the Hi-seq 2000 (Salk Institute), generating 2X 100 bp paired end reads. We aligned reads of 2 replicate WT ESC RNA-seq libraries and 3 replicate H1 TKO ESC RNA-seq libraries to the reference genome (NCBI37/mm9) with Tophat [39] and used Cufflinks and CuffDiff [40] for differential expression analysis of RNA-seq expression for a non-redundant collection of 20876 known RefSeq transcripts. We considered genes with a marginal p-value smaller than 0.05 and an absolute log2 fold-change bigger than 1 to be differentially expressed (598 genes).

Genome-wide DNA Methylation Analysis Using the HELP-tagging Assay

Genomic DNA was isolated, digested with HpaII and MspI, and Illumina library preparation was performed exactly as described previously [23]. Library sequencing was performed in the Epigenomics Core Facility at the Albert Einstein College of Medicine. We computed the HELP angle as described in [23] and used it as a measure for the percentage of methylated cytosines. We performed binomial tests for differential methylation and this resulted in 15492 differentially methylated sites with a p-value smaller than 1e-6.

DNaseI hypersensitivity assay

DNaseI hypersensitivity assay was essentially carried out as described in [41].In brief, nuclei were extracted in lysis buffer (15 mM Tris-HCl, 15 mM NaCl, 60 mM KCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM spermidine) by incubating for 10 minutes on ice. Then, nuclei were incubated for 3 min at 37°C in the same lysis buffer with 1 mM $CaCl_2$ and with limiting concentrations of the DNA endonuclease deoxyribonuclease I (DNase I). Reactions were stopped by adding stop buffer (50 mM Tris-HCl, 100 mM NaCl, 0.1% SDS, 100 mM EDTA, 1 mM spermidine, 0.5 spermine, pH8.0) purified fragments were recovered by sucrose ultracentrifugation, end-repaired and ligated with adapters, followed by sequencing on the Illumina sequencing platform. From an initial collection of 33 different DNase-seq libraries (17 WT, 16 TKO), we filtered out three high quality replicates in each condition according to their SPOT score. Reads were aligned to the reference genome (NCBI37/mm9) and we considered 36bp reads that aligned uniquely and contained no more than 2 mismatches as properly mapped reads. We used the Hotspot [42] algorithm to identify DNase I hypersensitive regions in all 6 six samples separately. For differential DNase-seq analysis between WT and TKO, we used the PoissonSeq R package [43]. We compared DNase-seq coverage in 89 875 different regions (with sufficient coverage in either condition)

and this resulted in a set of 4166 regions with statistically significant difference in coverage after applying a multiple testing procedure (false-discovery rate 5%).

Statistical analysis

All statistical analyses were performed under R/Bioconductor [44] using custom R scripts. Manipulation with and computation of statistics on genomic intervals and domains was done using the **GenomicRanges** package [45]. Analysis of ChIP-seq data and DNaseI hypersensitivity data and generation of heatmaps were done using the **compEpiTools** package (http://genomics.iit.it/groups/computational-epigenomics.html).

LIST OF ABBREVIATIONS

TAD: topologically associated domain
TKO: triple knock-out
DMR: differentially methylated region
ESC: embryonic stem cell
ICR: imprinting control region
DHS: Dnase 1 hypersensitivity site

COMPETING INTERESTS

The authors declare no competing interests.

AUTHORS' CONTRIBUTIONS

G.G., B.A.B. computationally analyzed the data, Y.Z., B.J.K., S.M.Y., M.J.A.M.V. and P.J.W. performed the experiments, T.S.M. generated and analyzed RNA-seq data, J.A.S. generated and analyzed Dnase I data, G.G., A.I.S. & WdL conceived the study, participated in its design and drafted the manuscript. All authors read and approved the final manuscript.

Description of additional data files

ACKNOWLEDGEMENTS

We thank Britta Bouwman for help with the figures, Michal Mokry and Edwin Cuppen for the help with high throughput sequencing, Stieneke van den Brink for ES cell culture, and Masako Suzuki and John Greally for guidance in preparing HELP-tagging libraries. This work was supported by National Institutes of Health Grant CA079057 to A.I.S. and an NWO/CW TOP grant (714.012.002), an NWO VICI grant 724.012.003, an EU grant 2010-259743 (MODHEP) and a European Research Council Starting Grant (209700, '4C') to W.d.L.

REFERENCES

1. Van Holde KE: *Chromatin*. New York, N.Y.: Springer; 1989.
2. Wolffe A: *Chromatin : structure and function.* 3rd ed. edn. San Diego, Calif. ; London: Academic; 1998.
3. Brown DT: *Histone variants: are they functionally heterogeneous?* Genome Biol 2001, **2:**REVIEWS0006.
4. Happel N, Doenecke D: *Histone H1 and its isoforms: contribution to chromatin structure and function.* Gene 2009, **431:**1-12.
5. Izzo A, Kamieniarz K, Schneider R: *The histone H1 family: specific members, specific functions?* Biol Chem 2008, **389:**333-343.
6. Sirotkin AM, Edelmann W, Cheng G, Klein-Szanto A, Kucherlapati R, Skoultchi AI: *Mice develop normally without the H1(0) linker histone.* Proc Natl Acad Sci U S A 1995, **92:**6434-6438.
7. Fan Y, Sirotkin A, Russell RG, Ayala J, Skoultchi AI: *Individual somatic H1 subtypes are dispensable for mouse development even in mice lacking the H1(0) replacement subtype.* Mol Cell Biol 2001, **21:**7933-7943.
8. Lin Q, Sirotkin A, Skoultchi AI: *Normal spermatogenesis in mice lacking the testis-specific linker histone H1t.* Mol Cell Biol 2000, **20:**2122-2128.
9. Drabent B, Saftig P, Bode C, Doenecke D: *Spermatogenesis proceeds normally in mice without linker histone H1t.* Histochem Cell Biol 2000, **113:**433-442.
10. Rabini S, Franke K, Saftig P, Bode C, Doenecke D, Drabent B: *Spermatogenesis in mice is not affected by histone H1.1 deficiency.* Exp Cell Res 2000, **255:**114-124.
11. Fantz DA, Hatfield WR, Horvath G, Kistler MK, Kistler WS: *Mice with a targeted disruption of the H1t gene are fertile and undergo normal changes in structural chromosomal proteins during spermiogenesis.* Biol Reprod 2001, **64:**425-431.
12. Fan Y, Nikitina T, Zhao J, Fleury TJ, Bhattacharyya R, Bouhassira EE, Stein A, Woodcock CL, Skoultchi AI: *Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation.* Cell 2005, **123:**1199-1212.
13. Yang SM, Kim BJ, Norwood Toro L, Skoultchi AI: *H1 linker histone promotes epigenetic silencing by regulating both DNA methylation and histone H3 methylation.* Proc Natl Acad Sci U S A 2013, **110:**1708-1713.
14. Maclean JA, Bettegowda A, Kim BJ, Lou CH, Yang SM, Bhardwaj A, Shanker S, Hu Z, Fan Y, Eckardt S, et al: *The rhox homeobox gene cluster is imprinted and selectively targeted for regulation by histone h1 and DNA methylation.* Mol Cell Biol 2011, **31:**1275-1287.
15. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B: *Topological domains in mammalian genomes identified by analysis of chromatin interactions.* Nature 2012, **485:**376-380.
16. Splinter E, Heath H, Kooren J, Palstra RJ, Klous P, Grosveld F, Galjart N, de Laat W: *CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus.* Genes Dev 2006, **20:**2349-2354.
17. Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T: *Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells.* Dev Cell 2006, **10:**105-116.
18. Melcer S, Hezroni H, Rand E, Nissim-Rafinia M, Skoultchi A, Stewart CL, Bustin M, Meshorer E: *Histone modifications and lamin A regulate chromatin protein dynamics in early embryonic stem cell differentiation.* Nat Commun 2012, **3:**910.
19. Woodcock CL, Skoultchi AI, Fan Y: *Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length.* Chromosome Res 2006, **14:**17-25.
20. Fan Y, Nikitina T, Morin-Kensicki EM, Zhao J, Magnuson TR, Woodcock CL, Skoultchi AI: *H1 linker histones are essential for mouse development and affect nucleosome spacing in vivo.* Mol Cell Biol 2003, **23:**4559-4572.
21. de Wit E, Bouwman BA, Zhu Y, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, et al: *The pluripotent genome in three dimensions is shaped around pluripotency factors.* Nature 2013, **501:**227-231.

22. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al: *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science 2009, **326:**289-293.
23. Suzuki M, Jing Q, Lia D, Pascual M, McLellan A, Greally JM: *Optimized design and data analysis of tag-based cytosine methylation assays.* Genome Biol 2010, **11:**R36.
24. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al: *Spatial partitioning of the regulatory landscape of the X-inactivation centre.* Nature 2012, **485:**381-385.
25. Ernst J, Kellis M: *ChromHMM: automating chromatin-state discovery and characterization.* Nat Methods 2012, **9:**215-216.
26. Schuettengruber B, Oded Elkayam N, Sexton T, Entrevan M, Stern S, Thomas A, Yaffe E, Parrinello H, Tanay A, Cavalli G: *Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila.* Cell Rep 2014, **9:**219-233.
27. Zhang Y, Liu Z, Medrzycki M, Cao K, Fan Y: *Reduction of Hox gene expression by histone H1 depletion.* PLoS One 2012, **7:**e38829.
28. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G: *Three-dimensional folding and functional organization principles of the Drosophila genome.* Cell 2012, **148:**458-472.
29. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL: *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell 2014, **159:**1665-1680.
30. Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J: *Spatial organization of the mouse genome and its role in recurrent chromosomal translocations.* Cell 2012, **148:**908-921.
31. Chandra T, Ewels PA, Schoenfelder S, Furlan-Magaril M, Wingett SW, Kirschner K, Thuret JY, Andrews S, Fraser P, Reik W: *Global Reorganization of the Nuclear Landscape in Senescent Cells.* Cell Rep 2015.
32. Noordermeer D, Branco MR, Splinter E, Klous P, van Ijcken W, Swagemakers S, Koutsourakis M, van der Spek P, Pombo A, de Laat W: *Transcription and chromatin organization of a housekeeping gene cluster containing an integrated beta-globin locus control region.* PLoS Genet 2008, **4:**e1000016.
33. Palstra RJ, Simonis M, Klous P, Brasset E, Eijkelkamp B, de Laat W: *Maintenance of long-range DNA interactions after inhibition of ongoing RNA polymerase II transcription.* PLoS One 2008, **3:**e1661.
34. Therizols P, Illingworth RS, Courilleau C, Boyle S, Wood AJ, Bickmore WA: *Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells.* Science 2014, **346:**1238-1242.
35. Splinter E, de Wit E, van de Werken HJ, Klous P, de Laat W: *Determining long-range chromatin interactions for selected genomic sites using 4C-seq technology: from fixation to computation.* Methods 2012, **58:**221-230.
36. Langmead B, Trapnell C, Pop M, Salzberg SL: *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol 2009, **10:**R25.
37. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS: *Model-based analysis of ChIP-Seq (MACS).* Genome Biol 2008, **9:**R137.
38. Boyle AP, Guinney J, Crawford GE, Furey TS: *F-Seq: a feature density estimator for high-throughput sequence tags.* Bioinformatics 2008, **24:**2537-2538.
39. Trapnell C, Pachter L, Salzberg SL: *TopHat: discovering splice junctions with RNA-Seq.* Bioinformatics 2009, **25:**1105-1111.
40. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nat Protoc 2012, **7:**562-578.

41. Vierstra J, Rynes E, Sandstrom R, Zhang M, Canfield T, Hansen RS, Stehling-Sun S, Sabo PJ, Byron R, Humbert R, et al: *Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution*. Science 2014, **346:**1007-1012.
42. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al: *The accessible chromatin landscape of the human genome.* Nature 2012, **489:**75-82.
43. Li J, Witten DM, Johnstone IM, Tibshirani R: *Normalization, testing, and false discovery rate estimation for RNA-sequencing data.* Biostatistics 2012, **13:**523-538.
44. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, et al: *Bioconductor: open software development for computational biology and bioinformatics.* Genome Biol 2004, **5:**R80.
45. Lawrence M, Huber W, Pages H, Aboyoun P, Carlson M, Gentleman R, Morgan MT, Carey VJ: *Software for computing and annotating genomic ranges.* PLoS Comput Biol 2013, **9:**e1003118.

SUPPLEMENTARY MATERIALS

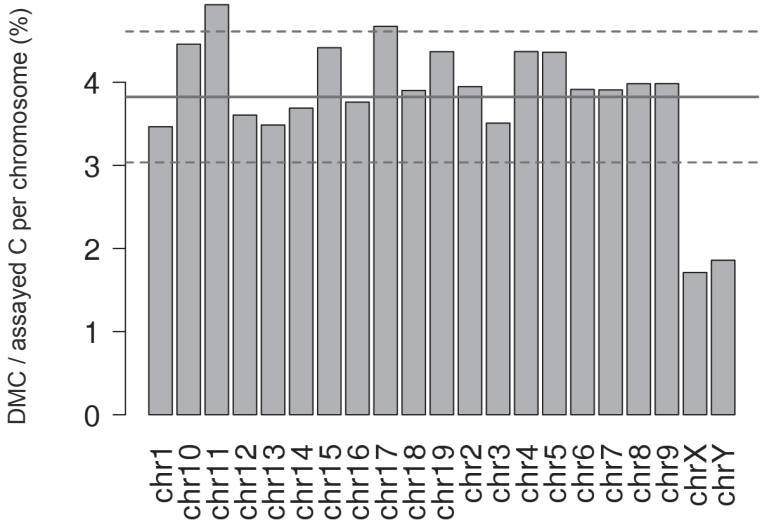


Figure S1. DNA methylation changes per chromosome in histone H1 depleted ES cells
Barplots showing the percentage of differentially methylated cytosine (DMC) per million of assayed cytosines for each chromosome. The red solid and dashed lines mark the mean and the points one standard-deviation away from the mean.

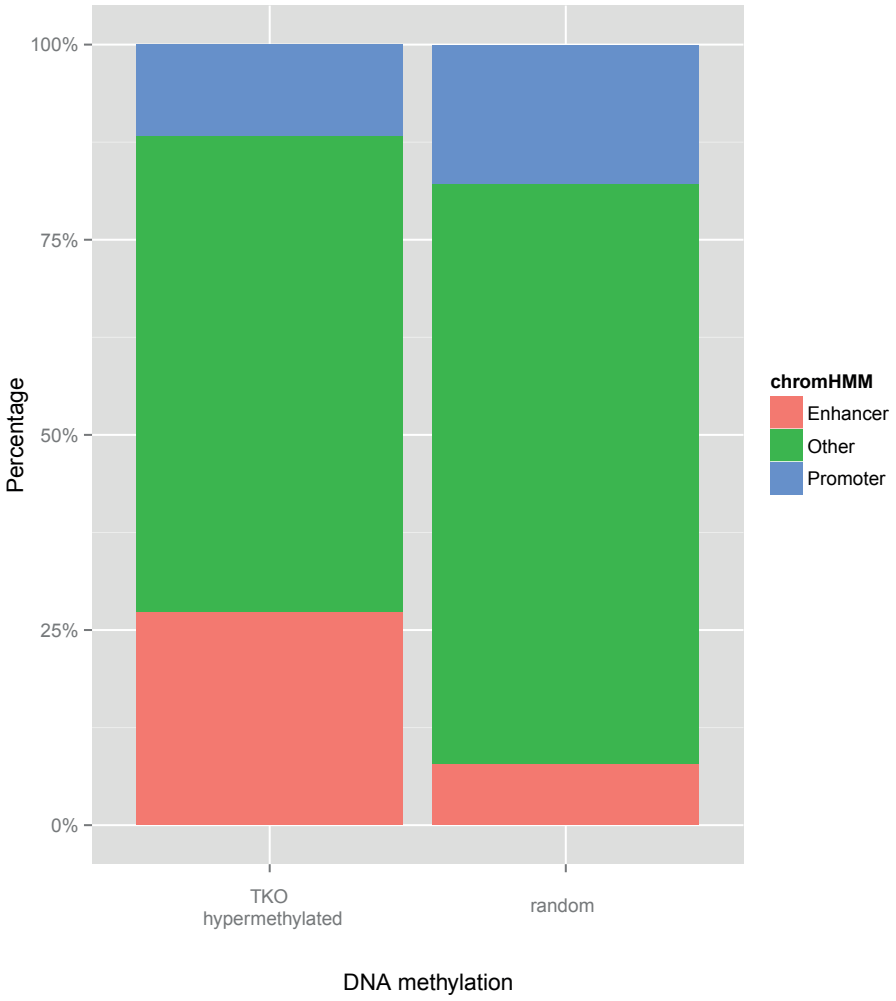


Figure S2. Hypermethylation of CpG in H1 TKO cells preferentially occurs at predicted enhancers
Spatial distribution of genomic sites where either hypo- or hypermethylation is observed in H1 depleted cells. We analyzed the location of differentially methylated sites with respect to different types of chromatin defined by the ChromHMM algorithm (based on a large collection of mouse ES cell ChIP-Seq data from the ENCODE consortium). For comparison, we include the same distribution for a random selection of sites where DNA methylation status was determined in the HELP-tagging experiment.

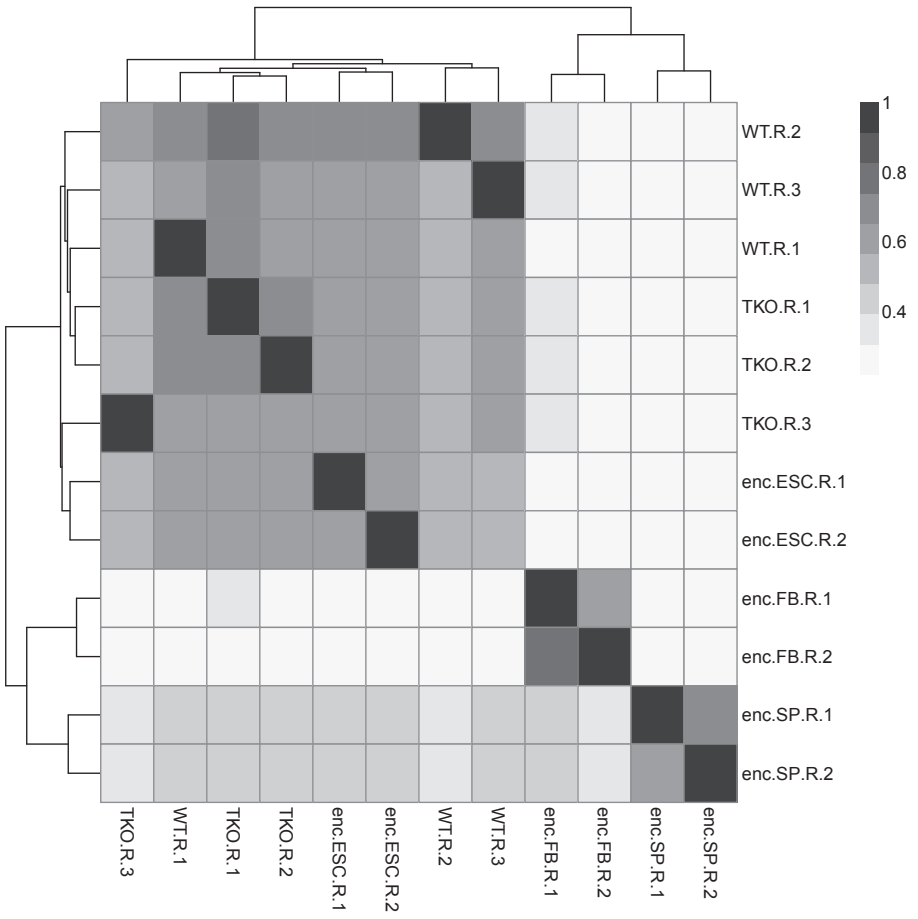


Figure S3. Very similar DHS profiles between Histone H1 TKO and wildtype ES cells
Clustered heatmap showing the similarity between different DNase I hypersensitivity (DHS) datasets in terms of genomic overlap of regions of enriched DNase-seq coverage. We compare our triplicate DHS experiments in H1 TKO and matched WT ES cells to published data from the ENCODE project (duplicate experiments in mouse ES cells, spleen (SP) and fibroblasts (FB)).

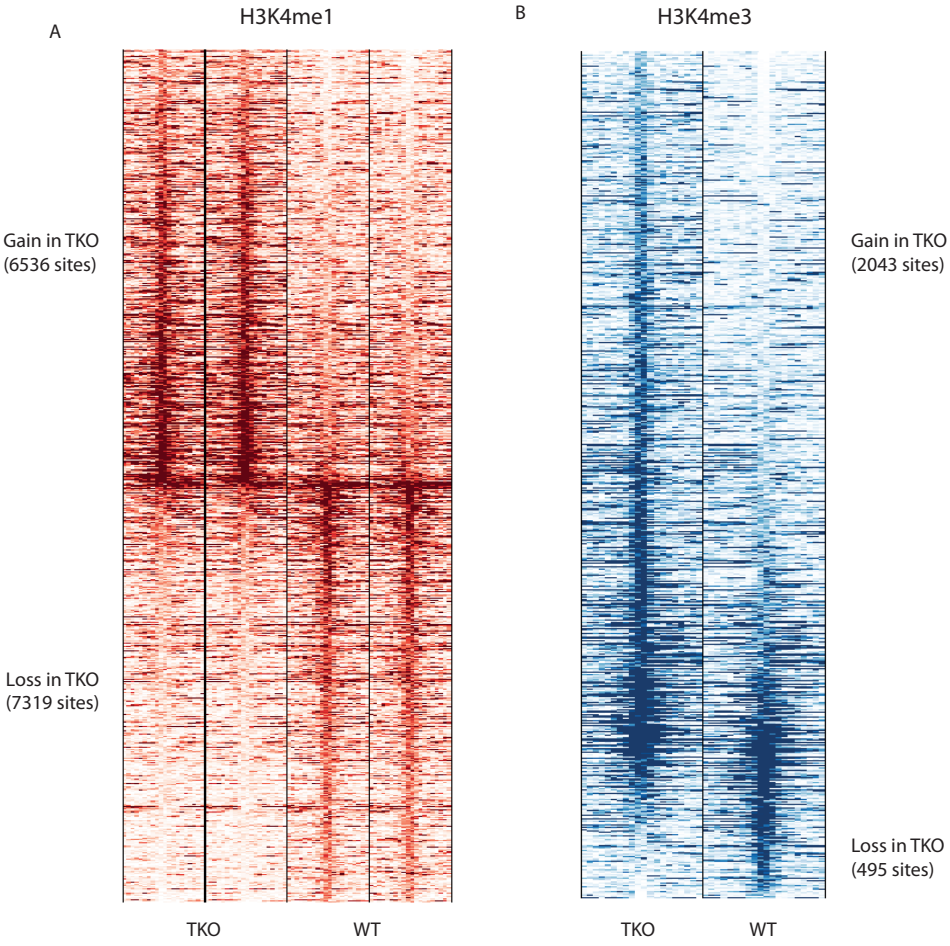


Figure S4. Large-scale changes in H3K4me1 and H3k4ME3 sites upon depletion of histone H1
Heatmaps of ChIP-seq enrichment for histone marks H3K4me1 and H3K4me3 in wildtype and H1 TKO cells. Genomic sites represented by rows in the heatmap are sites where significant changes in H3K4me3 enrichment are observed. Rows are ranked by the magnitude of that change from top to bottom

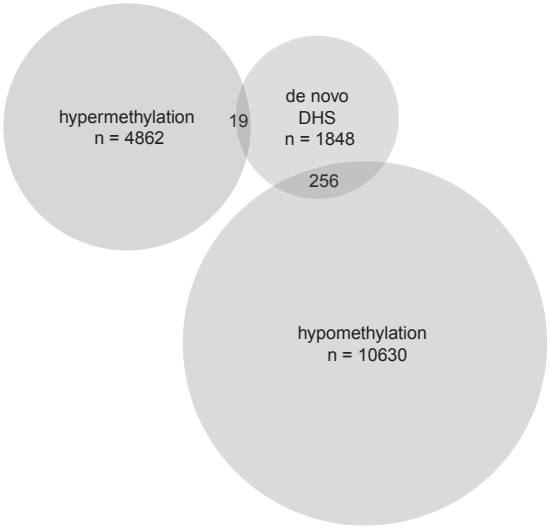


Figure S5. Over ten percent of the de novo formed DHS also show loss in CpG methylation in H1 TKO cells

Venn diagram showing counts of de novo formed DHS and their overlap with sites that show a significant gain or loss of DNA methylation in TKO.

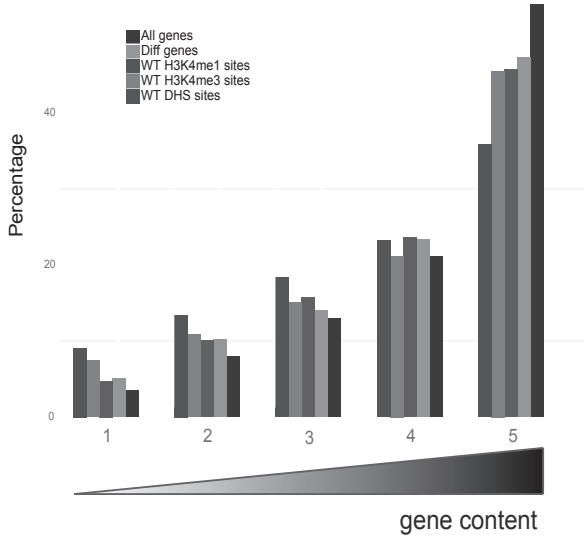


Figure S6. Active chromatin marks accumulate in the most gene-dense TADs

Barplots showing the percentages of differentially expressed genes compared to all genes in groups of TADs ranked according to the number of genes contained within a TAD. Also shown is the distribution of sites enriched for H3K4me1, H3K4me3, DNaseI hypersensitive sites in WT ES cells over these 5 categories of TADs.

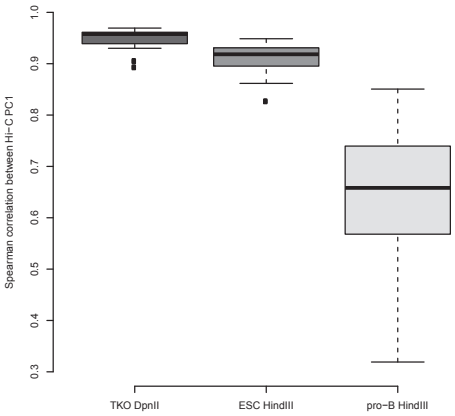


Figure S7. Higher-prder genome topology is very similar between wildtype and histone H1 TKO cells
The coefficients of the first principal component of the normalized Hi-C interaction matrix can be used to distinguish between the A and B chromatin compartments (see Figure 5A). By comparing the correlation of these coefficients (per chromosome) between conditions, we can measure how similar they are in terms of A and B compartmentalization. Figure S7 shows boxplots comparing the correlation (per chromosome) of PC1 coefficients between our WT DpnII Hi-C map, and the Hi-C maps indicated on the x-axis, i.e. TKO DnpII, ESC HindIII and pro-B HindIII. This shows that in terms of A and B compartment organization, the TKO Hi-C map is more similar to the WT Hi-C map than to a previously published mouse ES Hi-C map (small difference) and to a Hi-C map published in a completely different cell type (large difference).

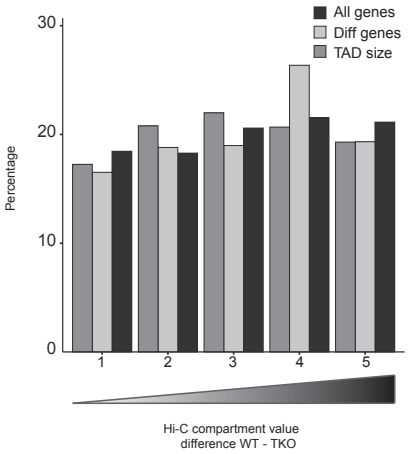


Figure S8. Changes in compartment organization are not related to gene content, TAD size or differential expression
Barplots showing the percentages of genes in 5 equally sized groups of TADs ranked according to the difference in Hi-C compartment value between WT and H1 TKO ES cells. Also shown is the distribution of differentially expressed genes over these 5 groups of TADs. The total genomic size of the TADs in each group as a percentage of the total genomic size of all TADs is plotted as a reference.

5

Discussion

The work described in this thesis aimed to understand the roles of chromatin proteins and transcription factors in shaping the 3D organization of the genome. Certain proteins have been shown to be critical for local loop formation. Examples of these include the erythroid-specific transcription factors KLF1 and GATA1 [1, 2]. Our focus was more on the global, higher-order, 3D organization of the genome. In our analyses we included pluripotency factors, transcriptional repressors and linker histone H1. We used site-specifically integrated Lac operator repeat sequences to site-specifically recruit proteins of interest. In addition we studied chromosome conformation features in the absence of pluripotency transcription factors and in cells with reduced histone H1. We demonstrated that tethering of most factors had little impact on the contact spectrum of our two target sites. However, certain factors did cause the target sequences to adopt new positions in the nucleus. In often cases they stimulated the establishment of long-range chromosomal contacts with sequences bound these same factors. This observation is in agreement with the previous proposed dog-on–a-leash model [3, 4]. This model suggests that the position of the great majority of individual genes is mainly defined by the overall chromosome topology. Genes themselves and regulatory elements have very limited freedom to move around and essentially sample a predefined genomic environment for preferred contacts with similarly typed chromatin regions. Only some loci, usually large blocks of repetitive sequences, are dominant over this type of organization and autonomously determine their 3D genomic environment.

In two instances we found that a locus can adopt completely new contacting partners, once upon recruitment of Ezh2 and in the other case after Suv39h1 association. This, however, was genomic context dependent, since only one of the two loci present in the same cell showed such dramatic repositioning. This locus chose its new contacting partners based on which protein was recruited, which was surprising given that the 3D genome is thought to be subdivided into two types of nuclear compartments, the A and B compartments. Recently however, a more detailed chromatin contact map revealed that the genome can be further subdivided into six smaller compartments, with participating chromosomal regions having a median length of 185kb (Table 1). In such compartments, loci tend to possess similar features including epigenetic modifications and long-range interaction patterns [5]. The existence of multiple types of active and inactive compartments that separate according to their epigenetic make-up is very much in agreement with our observation that depending on the proteins recruited, there are multiple compartments a locus can be dragged to. We found that locus repositioning was uncoupled from transcriptional changes. This observation is consistent with a recently published finding that chromatin de-condensation at several loci

Table 1. The development of Hi-C with corresponding proposed models

YEAR	VALID READS	RESOLUTION	MODEL
2009	6,7million	1MB	A/B compartments (5~20MB mega domain)
2012	1,7 billion	5kb	TAD (1MB domain)
2014	4,9 billion	1kb	six compartments (median size 185kb domain)

is able to relocate genes without any effect on their expression [6]. This again may suggest that remodeling of chromatin, rather than transcription, determines the position of a locus.

Previously, others have documented that protein recruitment to LacO arrays can impact on chromatin organization. In one instance large-scale directed chromatin movement over several micrometers was observed in interphase nuclei, after recruitment of the transcriptional activator VP16 to a Lac operator array that was 10-20x larger than our array [7, 8]. This so far is the exception though as in other cases repositioning was found to only occur after mitosis. Recruitment for example of components of the nuclear lamina to specific gene loci did result in their repositioning to the nuclear lamina but only after cell division. Also, the separation of chromatin between the nuclear interior and nuclear periphery appears quite stable during interphase but can drastically be re-organized during and after mitosis [9]. These observations are in line with the theory that the overall chromatin structure is defined in early G1, when chromatin is loose and mobile. At this stage, preferred contacts can be established to constrain the reminder of the surrounding region. Some chromatin segments, such as rDNA clusters and other highly repetitive sequences are thought to have a dominant role in chromosome positioning as they cluster to form nucleoli and chromocenters in nearly every cell. Their positioning will obviously influence that of other sequences linked on the same chromosome [4, 10]. Once the genome of a given cell has adopted its unique 3D configuration, movements of individual genes during the rest of interphase is generally constrained to a small nuclear subvolume, as seen by live cell imaging studies [11-13].

How protein factors contribute to the 3D genome folding?

During mitosis, chromatin undergoes dramatic changes with dis-associating many transcription factors. Some DNA-binding proteins could persist through cell divisions to facilitate their regulation of target genes [14-16]. PcG proteins were found to retain at chromatin domain borders during mitosis, which may help to re-establish the genomic structure [17]. It may also explain why PcG protein binding sites often cluster together in interphase nuclei. A recent 4C study confirmed that on a genome-wide scale, PcG protein binding sites tend to co-localize with each other. These specific contacts could be destroyed by depleting one of the PcG proteins that caused complete loss of H3K27me3; the rest of the global organization remained intact [18]. This is consistent with what we found in Chapter 3. Pluripotency factor binding sites show a preference in spatial clustering, which may facilitate or fine-tune the regulation of the genes they control. Removing a pluripotency factor or targeting it to certain loci only changed specific contacts, while the overall genome-wide contact maps remained similar (Fig.1) [19].

The functional significance of long-range DNA contacts

What about the functional relevance of long-range DNA contacts within and between chromosomes? As mentioned, individual genes are thought to sample a relatively small nuclear subvolume once the overall genome topology is set early in G1. Here, they can get engaged in intra- or inter-chromosomal contacts. Indeed, interchromosomal contacts have also been reported before [20-22]. In naive $CD4^+$ helper cells, the regulatory regions of TH2 cytokine genes on chromosome 11 were found in proximity with the promoter

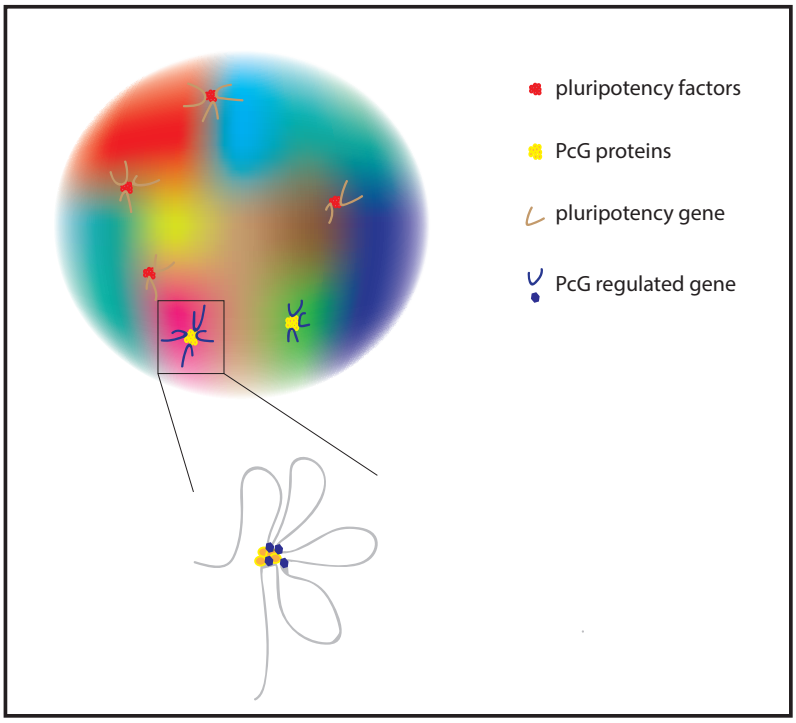


Figure 1. Specialized protein factors physically remodel chromatin structure and their responsive genes for their regulation. Genes, from the same or different chromosome territories (colored areas), controlled by the same factor prefer to meet each other in the nuclear space, forming long-range chromatin loops.

region of the IFN-gamma gene on chromosome 10. This cell type specific contact was suggested to poise the genes for activation later in differentiation [23]. Another study on the olfactory genes proposed that a single enhancer could stochastically contact any one of the ~1200 olfactory receptor genes dispersed across the mouse genome, thereby activating that receptor. However, deletion of this enhancer only affected the expression of olfactory receptor genes on the same chromosome, indicating that other mechanisms may also contribute in this process [24]. More recently, Fanucchi *et al* used TALENs to disrupt intra- or inter-chromosomal contacts in a multigene complex. Cleavage of a sequence was claimed to disrupt the interchromosomal contacts and appeared to affect the expression of interacting genes. From this work it was reported that long-range chromosomal contacts are functionally important for gene regulation [25].

Cell-specific 3D genome

Whether or not these rare claims on functional interchromosomal interactions are valid or relevant, remains to be seen. From an evolutionary standpoint it seems unlikely that

functionally important DNA sequences are positioned on other chromosomes than their target genes. Also, the stochastic folding of the genome in each cell and the inability of individual genes or regulatory sequences to search the nuclear interior for preferred genomic contacts makes it difficult to conceptualize how interchromosomal contacts can control gene expression at the cell population level. In this respect the observations made by Noordermeer *et al* appear relevant. They characterized mice with a human b-globin LCR integrated in chromosome 8 [3]. This ectopic super enhancer changed the chromatin structure and the expression of genes locally around the integration site, as well as the expression of the endogenous mouse b-globin gene on chromosome 7. This gene however was only activated in the few cells where the human LCR on chromosome 8 happened to be in close proximity with the gene. This experiment proved genetically that long-range chromatin interactions can be functional important for gene regulation, but strongly suggested that this will only be happening in the rare cells forming this contact. Interchromosomal DNA interactions and cell-specific 3D genomes can therefore contribute to cell-to-cell variation in gene expression. The current need is to combine gene expression analysis and chromatin structure profiling at the single cell level to better understand the functional relevance of specific long-range DNA contacts. With the development of more advanced microscopy and single-molecule techniques, it may become possible to follow a gene's contacts with enhancers and/or other sequences and quantify its expression in single cells. This may uncover the quantitative relationship between local and distal DNA contacts and the transcriptional output of a gene.

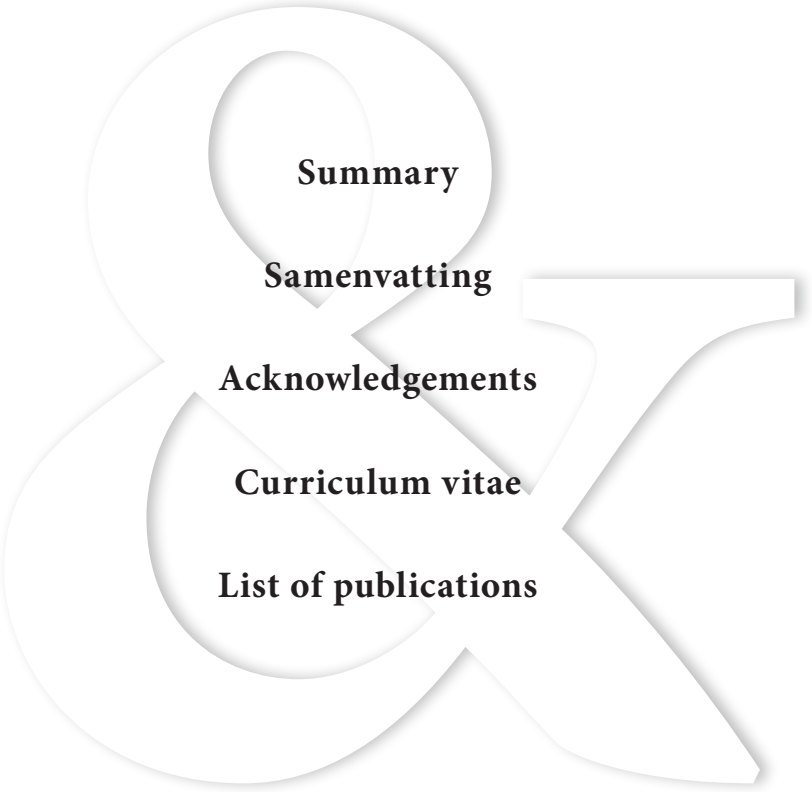
How dynamic are long-range DNA contacts?

A final interesting follow-up issue building on the toolbox described in this thesis would be to use the LacO targeted cells to follow the stability and dynamics of long-range contacts during interphase and over cell division. For this, a second set of operator sequences needs to be targeted in cis (on the same chromosome) and in trans (on other chromosomes) in cells carrying the already integrated LacO cassettes. While in interphase two loci are expected to show a relatively stable spatial separation that may well be different between individual cells, the interesting question is whether there is memory in their relative positioning over cell division. Does the genome re-establish specific interchromosomal contacts in daughter cells? And is there memory in the overall folding of a chromosome over cell division? Demonstrating memory in long-range DNA contacts would strongly support the concept that they may be of functional significance for individual cells and their offspring.

REFERENCES

1. Drissen, R., et al., *The active spatial organization of the beta-globin locus requires the transcription factor EKLF.* Genes Dev, 2004. **18**(20): p. 2485-90.
2. Vakoc, C.R., et al., *Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1.* Mol Cell, 2005. **17**(3): p. 453-62.
3. Noordermeer, D., et al., *Variegated gene expression caused by cell-specific long-range DNA interactions.* Nat Cell Biol, 2011. **13**(8): p. 944-51.

4. Krijger, P.H. and W. de Laat, *Identical cells with different 3D genomes; cause and consequences?* Curr Opin Genet Dev, 2013. **23**(2): p. 191-6.
5. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell, 2014. **159**(7): p. 1665-80.
6. Therizols, P., et al., *Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells.* Science, 2014. **346**(6214): p. 1238-42.
7. Chuang, C.H., et al., *Long-range directional movement of an interphase chromosome site.* Curr Biol, 2006. **16**(8): p. 825-31.
8. Khanna, N., Y. Hu, and A.S. Belmont, *HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation.* Curr Biol, 2014. **24**(10): p. 1138-44.
9. Kind, J., et al., *Single-cell dynamics of genome-nuclear lamina interactions.* Cell, 2013. **153**(1): p. 178-92.
10. Politz, J.C., D. Scalzo, and M. Groudine, *Something silent this way forms: the functional organization of the repressive nuclear compartment.* Annu Rev Cell Dev Biol, 2013. **29**: p. 241-70.
11. Marshall, W.F., et al., *Interphase chromosomes undergo constrained diffusional motion in living cells.* Curr Biol, 1997. **7**(12): p. 930-9.
12. Heun, P., et al., *Chromosome dynamics in the yeast interphase nucleus.* Science, 2001. **294**(5549): p. 2181-6.
13. Chubb, J.R., et al., *Chromatin motion is constrained by association with nuclear compartments in human cells.* Curr Biol, 2002. **12**(6): p. 439-45.
14. Xing, H., et al., *Mechanism of hsp70i gene bookmarking.* Science, 2005. **307**(5708): p. 421-3.
15. Delcuve, G.P., S. He, and J.R. Davie, *Mitotic partitioning of transcription factors.* J Cell Biochem, 2008. **105**(1): p. 1-8.
16. Xing, H., N.L. Vanderford, and K.D. Sarge, *The TBP-PP2A mitotic complex bookmarks genes by preventing condensin action.* Nat Cell Biol, 2008. **10**(11): p. 1318-23.
17. Follmer, N.E., A.H. Wani, and N.J. Francis, *A polycomb group protein is retained at specific sites on chromatin in mitosis.* PLoS Genet, 2012. **8**(12): p. e1003135.
18. Denholtz, M., et al., *Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and polycomb proteins in genome organization.* Cell Stem Cell, 2013. **13**(5): p. 602-16.
19. de Wit, E., et al., *The pluripotent genome in three dimensions is shaped around pluripotency factors.* Nature, 2013. **501**(7466): p. 227-31.
20. Apostolou, E. and D. Thanos, *Virus Infection Induces NF-kappaB-dependent interchromosomal associations mediating monoallelic IFN-beta gene expression.* Cell, 2008. **134**(1): p. 85-96.
21. Ling, J.Q., et al., *CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1.* Science, 2006. **312**(5771): p. 269-72.
22. Zhao, Z., et al., *Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions.* Nat Genet, 2006. **38**(11): p. 1341-7.
23. Spilianakis, C.G., et al., *Interchromosomal associations between alternatively expressed loci.* Nature, 2005. **435**(7042): p. 637-45.
24. Lomvardas, S., et al., *Interchromosomal interactions and olfactory receptor choice.* Cell, 2006. **126**(2): p. 403-13.
25. Fanucchi, S., et al., *Chromosomal contact permits transcription between coregulated genes.* Cell, 2013. **155**(3): p. 606-20.



Summary

Samenvatting

Acknowledgements

Curriculum vitae

List of publications

SUMMARY

Cells are the basic building units of our body. Each tiny cell nucleus contains roughly 2 meters of DNA, if completely unwound. DNA is packed into chromatin to make it fit in such a small cellular space and to fulfill its function. Chromatin structure has been studied for a long time. The basic unit of chromatin is the nucleosome, composed of core histones with DNA wrapped around it. The nucleosomes form a "beads on a string" structure, with linker DNA connecting two adjacent nucleosomes. The nucleosome array is further packed and folded into higher order structures. During interphase, the spatial organization of chromatin is not random, and genome architecture is considered to be important for its functioning. Chromosomes each occupy a distinct nuclear space, known as chromosome territories. They organize themselves in a probabilistic manner, with larger and/or inactive chromosomes preferentially adopting a peripheral position and smaller and/or active chromosomes often being more in the nuclear interior. Similar probabilistic rules underlie the positioning of chromosomal domains, also called topologically associated domains, or TADs. Active TADs and inactive TADs each tend to cluster among each other, with the active one being more often inside the cell nucleus and the inactive ones being at the periphery. Between cells and during cellular division however the exact genomic neighborhood (the spatially neighboring TADs) of each given TAD is likely to be different.

Microscopy studies and high throughput techniques such Hi-C, 4C but also Dam-ID have greatly increased our knowledge of the 3D structure of the genome. Initial Hi-C data revealed that chromatin is partitioned into two genome-wide compartments: the A and B compartment. The A compartment contains chromosomal regions with open chromatin, active genes and many DHSs, while the B compartment mainly includes gene deserts with inactive histone modifications, lacking DHSs. As the sequencing depth increased, those A and B compartments were subdivided into 6 smaller compartments, each containing TADs with a shared chromatin signature. In TADs, local contacts occur much more frequently than between domains. Surprisingly, TAD organization is stable across different cell types and highly conserved over evolution. The mechanisms defining TADs and TAD positioning are still unclear. Insulator binding protein CTCF, housekeeping genes, SINEs and tRNAs are enriched at TAD boundaries, which may function as a barrier. A major aim of my work has been to study the cause and consequences of TAD repositioning and unravel the factors that shape the 3D genome.

In chapter2, we have recruited different types of repressive chromatin regulator, including the pluripotency transcription factor Nanog, polycomb group protein Ezh2 and Suv39h1, a constitutive heterochromatin histone methytransferase factor, to two lacO/lacR recruitment platforms integrated in transcriptionally active genomic loci in the same cell. 4C-seq was applied to explore the consequences of protein recruitment on the spatial location of the LacO binding platform. We found that genomic loci differ in their ability to take up new positions in the nucleus: some sites are resistant to repositioning irrespective of the protein recruited to the platform, while others adopt new positions in a manner that depends on the protein recruited. In instances where a given locus shows repositioning, folding of the remainder of the chromosome stays largely intact though. This is consistent

with the recently proposed "dog-on-a-lead" model that implies that the overall folding of a chromosome will impose restraints on the mobility and capacity of a locus to look for preferred interaction partners. Invariably, the trans-acting factors that induced repositioning all positioned the LacO cassette in a genomic neighborhood that had a chromatin signature corresponding to the factor of interest. Thus, recruitment of Suv39h1, which induces H3K9 trimethylation, dragged the LacO cassette and its subTAD to other genomic loci with high levels of H3K9me3 and low transcriptional activity. Similarly, recruitment of Ezh2, which deposits trimethylation on H3K27, brought the cassette in close spatial proximity to the HoxB and a Cbx gene cluster, both rich in this same histone mark. In both cases repositioning had no consequences on the expression of endogenous genes immediately surrounding the lacO site. This observation fits with a recent published study that locus decondensation instead of transcription was able to reposition endogenous genes from the nuclear periphery to the nuclear interior. Collectively, this suggests that nuclear organization is not a major determinant of gene expression, but may be important for maintaining a given transcriptional or chromatin state. Based on these findings, we conclude that gene expression and high-order chromosome topology beyond TADs are not causally related. Instead, they seem to be independently controlled by locally associated trans-acting factors.

Chapter 3 focuses on uncovering the principles of chromatin structure in pluripotent stem cell. In pluripotent stem cell nuclei, chromatin is relatively open and structure is less defined, hyperdynamic and permissive for regulators to bind. We combined 4C and Hi-C data with chromatin factor binding data to demonstrate that inactive chromatin is less organized in pluripotent nuclei. Additionally, we find that enhancers show tissue-specific contacts. Core pluripotency factors like Nanog, Sox2, Oct4 and Klf2 regulate a large number of genes and are essential factors to keep the pluripotent state. We found clusters of Nanog-, Oct4-, Sox2- binding sites showing a strong preference to interact with each other. This unique feature is only found in pluripotent nuclei, and depends on the presence of pluripotency factors. Artificially targeting Nanog to an unrelated genomic site increased its contact frequency with other Nanog binding sites on the same chromosome, suggesting that Nanog has a direct role in this specific spatial clustering. It is likely that pluripotency factors collaborate with each other to remodel the chromatin structure and the positioning of targeting genes, which in turn may facilitate the robust control of target gene expression and help keeping the pluripotent state. We also found that targeting or removing a pluripotency factor from the genome only changes specific genomic contacts, while the over structure conformation remains intact. This observation again supports the "dog-on-a-lead" model.

In chapter 4, we investigated the role of linker histone H1 in the 3D organization of the genome. Histone H1 binds to nucleosome core particles near the DNA entry/exit position and to the linker DNA between core particles, stabilizing the association of core particles and DNA and facilitating folding of oligonucleosome arrays into compact structures. H1-depleted ES cells with a 50% reduction in the ratio of H1 to the core particles showed decreased local chromatin compaction and selective changes in gene expression. Here we used more genome-wide approaches to study the effect of H1 depletion on chromatin conformation and function. Our data showed that cells require normal histone H1 levels

to maintain a proper regulatory landscape. Genome-wide analysis of DNA methylation, histone modifications and DNaseI hypersensitivity profiling revealed that thousands of potential regulatory sites change their epigenetic signature upon 50% reduction of H1. Hi-C data showed that chromatin structure largely remained intact with reduced levels of histone H1. However, in histone H1 depleted cells genomic loci more easily engaged in long range chromatin contacts, indicating that histone H1 may have a function in topological segmentation of chromosomes. Those TAD configuration changes coincide with epigenetic landscape changes but not with transcriptional output changes, supporting the concept we raised in chapter 2 that transcriptional control and nuclear positioning of TADs are not causally related but independently controlled by the locally associated trans-acting factors.

SAMENVATTING

Cellen zijn de basis bouwstenen van ons lichaam. Iedere celkern bevat zo'n 2 meter DNA, gemeten in compleet uitgevouwen staat. DNA is verpakt in chromatine om het te laten passen in de kleine celkern en om het zijn functie te vervullen. De structuur van chromatine is al lange tijd een veel onderzocht onderwerp. Chromatine bestaat uit basis eenheidjes, nucleosomen, welke bestaan uit een kern van histon eiwitten waar het DNA omheen is gevouwen. De nucleosomen vormen een structuur als "kralen aan een snoer", waarbij het DNA tussen twee naburige nucleosomen linker DNA wordt genoemd. Het snoer van nucleosomen is vervolgens verder opgevouwen tot hogere order chromatine structuren. De ruimtelijke verdeling van chromatine tijdens interfase is niet willekeurig, en de organisatie en architectuur van het genoom worden beschouwd als belangrijke factoren voor het functioneren. Ieder chromosoom neemt een eigen ruimte in in de celkern, ook wel chromosoom territorium wordt genoemd. Chromosomen organiseren zichzelf op een probabilistische wijze, waarbij grotere en/of inactieve chromosomen een perifere positie aannemen, terwijl kleinere en/of actieve chromosomen zich vaak meer centraal in de kern bevinden. Soortgelijke probabilistische positionering vindt plaats voor chromosomale domeinen, ook wel topologisch-geassocieerde-domeinen genaamd, of TADs. Actieve TADs en inactieve TADs zijn vaak als clusters te vinden, waarbij de actieve TADs met name in het binnenste van de celkern worden aangetroffen, terwijl de inactieve TADs meer naar de buitenkant positioneren. De exacte genomische omgeving van een bepaalde TAD in de celkern verschillen waarschijnlijk tussen cellen en over opeenvolgende celdelingen.

Onze kennis van de 3D structuur van het genoom is sterk vergroot door onderzoek met microscopie en high-throughput technieken zoals Hi-C, 4C, maar ook Dam-ID. Initiële HiC data lieten zien dat chromatine is verdeeld in twee genoom-wijde compartimenten: het A en het B compartiment. Compartement A bevat chromosomale gebieden die bestaan uit open chromatine, actieve genen en veel DHSs, en het B compartiment bestaat met name uit grote gebieden zonder genen, bekleedt met inactieve histon modificaties, en zonder DHSs. HiC experimenten met grotere sequence diepte hebben geleid tot het onderverdelen van de A en B compartimenten in 6 kleinere compartimenten welke ieder TADs bevatten met overeenkomstige chromatine eigenschappen. De contact frequentie binnen TADs is veel hoger dan tussen TADs. Het is opvallend dat deze TAD-organisatie sterk geconserveerd is in verschillende cel typen en tussen soorten. De mechanismes die TAD vorming en positionering definiëren zijn momenteel nog niet bekend. De sequenties die TADs begrenzen zijn verreikt voor huishoud-genen, SINEs, tRNAs, en de binding van het insulator-bindend eiwit CTCF. Deze sequenties functioneren mogelijk als versperringen of barrières. In mijn werk lag een sterke focus op het onderzoeken van de oorzaak en de consequenties van herpositionering van TADs en op het onderzoeken van de factoren die de 3D vouwing van het genoom vormgeven.

In hoofdstuk 2 hebben we verschillende represserende chromatine regulatoren naar twee lacO/lacR platformen gerekruteerd, welke waren geïntegreerd in transcriptioneel actieve genomische loci in dezelfde cel. De chromatine regulatoren die we hier laten zien zijn pluripotency transcriptie factoren Nanog, polycomb groep eiwit Ezh2 en Suv29h1 en

een constitutief heterochromatine histone methyl transferase factor. Om de consequenties van deze targeting op de 3D lokalisatie van het lacO bindings platform te onderzoeken werd na de rekrutering van ieder van deze eiwitten 4C-seq uitgevoerd. Dit liet zien dat er verschillen bestaan tussen de mate waarin verschillende genomische loci nieuwe posities kunnen aannemen in de kern: sommige plekken bleken resistent tegen repositionering, onafhankelijk van het eiwit dat naar het platform werd gerekruteerd, terwijl andere loci nieuwe posities aannamen afhankelijk van het type eiwit. In de gevallen van locus repositionering bleek dat de vouwing van de rest van het chromosoom grotendeels intact bleef, wat overeen komt met het recent voorgestelde 'dog-on-a-lead', ofwel hond-aan-een-lijn, model. Dit model suggereert dat de globale vouwing van een chromosoom een belemmering vormt voor de mobiliteit van een individueel locus en daarbij voor de capaciteit van dit locus om op zoek te gaan naar andere loci voor preferentiële 3D interacties. Zonder uitzondering zagen wij dat de trans-acting factoren die erin slaagden de lacO cassette te repositioneren zorgden voor verplaatsing naar een genomische omgeving met chromatine eigenschappen die overeenkwamen met de gerecruteerde factor. Zo zorgde de rekrutering van Suv29h1, een factor die H3K9 trimethylatie induceert, ervoor dat de lacO cassette en zijn flankerende sub-TAD werd verplaatst naar andere genomische plekken die gemarkeerd waren met veel H3K9me3 en weinig transcriptionele activiteit. Op een zelfde wijze laten we zien dat Ezh2, verantwoordelijk voor de trimethylering van H3K27, in staat bleek de cassette te repositioneren naar de HoxB en Cbx gen clusters, welke beiden verrijkt zijn voor deze histon modificatie. In beide gevallen werd geen effect waargenomen op de expressie van de endogene genen die direct naast de lacO integratie plek liggen. Deze observatie strookt met een recentelijk gepubliceerd onderzoek waarin de decondensering van een locus genoeg bleek te zijn om endogene genen te repositioneren van de periferie naar het midden van de celkern, en dat hier geen transcriptionele activatie voor nodig was. Deze gecombineerde observaties duiden er op dat nucleaire organisatie geen grote beslissende rol speelt in gen expressie, maar dat het belangrijk kan zijn voor het behouden van een bepaalde transcriptionele of chromatine staat. Op basis van onze resultaten concluderen we dat gen expressie en hogere-order chromosoom vouwing op het TAD-overstijgende niveau geen direct oorzakelijk verband vertonen. In plaats daarvan lijken ze onafhankelijk gecontroleerd te worden door lokaal geassocieerde trans-acting factoren.

Hoofdstuk 3 beschrijft de principes van chromatine structuur in pluripotente stamcellen. Het chromatine van pluripotente stamcellen is relatief open, heeft een minder duidelijke structuur en is hyperdynamisch en permissief voor de binding van regulerende factoren. Door middel van een combinatie van 4C en Hi-C data laten we hier zien dat inactief chromatine relatief ongeorganiseerd is in pluripotente kernen. Verder vonden we dat enhancers cell-specifieke contacten vormen. Pluripotency factoren Nanog, Sox2, Oct4 en Klf2 zijn essentieel voor het behoud van pluripotency en zijn betrokken bij de regulatie van een groot aantal genen. We vonden hier dat clusters van Nanog-, Oct4, en Sox2- bindingsplaatsen een sterke neiging hebben om elkaar te contacteren in de celkern. Dit verschijnsel werd alleen aangetroffen in pluripotente cellen op een wijze die afhankelijk is van de aanwezigheid van de pluripotency transcriptie factoren. Artificiële targeting van

Nanog naar een niet-gerelateerde plek in het genoom zorgde ervoor dat deze plek nieuwe contacten maakt met andere Nanog bindingsplekken op het zelfde chromosoom. Deze bevinding suggereert dat Nanog een directe rol heeft bij het tot stand brengen van deze specifieke ruimtelijke clustering. We denken dat het aannemelijk is dat pluripotency factoren samenwerken om chromatine structuur te remodeleren en genen te positioneren, waardoor mogelijk een robuuste controle over target gen expressie wordt bewerkstelligd die helpt bij het behouden van de pluripotente staat. Verder laten we zien dat het targeten of verwijderen van een pluripotency factor uit de kern slechts leidt tot het veranderen van een aantal specifieke genomische contacten, terwijl de globale chromosoom structuur intact blijft; een observatie die ook het 'dog-on-a-lead' model steunt.

In hoofdstuk 4 hebben we de rol van linker histone H1 in de 3D organisatie van het genoom onderzocht. Histone H1 bindt het nucleosoom dichtbij de DNA ingang/uitgang positie en het linker DNA tussen twee nucleosomen, waardoor het de associatie tussen nucleosomen en DNA stabiliseert en vouwing van oligo-nucleosoom arrays tot compacte structuren mogelijk maakt. Stamcellen gedepleteerd voor H1 hebben gemiddeld een 50% afname in de H1-tot-nucleosoom ratio, resulterend in een verminderde lokale chromatine compactheid en specifieke veranderingen in genexpressie. In dit hoofdstuk passen we genoom-brede technieken toe om het effect van H1 depletie op chromatine architectuur en functie te onderzoeken. Ons werk laat zien dat cellen een normaal histon H1 niveau nodig hebben om een goed functionerend regulatoir landschap te behouden. Genoom-brede analyse van DNA methylatie, histon modificatie, en DNase I hypersensitivity profilering lieten zien dat duizenden potentiële regulatoire gebieden hun epigenetische identiteit veranderen in de 50%-H1-gedepleteerde cellen. Door middel van Hi-C vonden we dat de chromatine structuur grotendeels intact blijft in deze cellen, maar loci lijken meer lange-afstands contacten aan te gaan, wat een indicatie is dat histon H1 een rol speelt in chromosoom segmentatie. Deze veranderingen in TAD configuratie correleren met veranderingen in het epigenetische landschap, terwijl er geen veranderingen zijn in de transcriptionele output. Onze bevindingen ondersteunen het concept dat we al benoemden in hoofdstuk 2, waarin we beschreven dat transcriptie regulatie en relatieve positionering van TADs geen direct oorzakelijk verband vertonen, maar dat ze onafhankelijk worden gecontroleerd door middel van lokaal geassocieerde trans-acting factoren.

ACKNOWLEDGEMENTS

Finally, this is the book! Looking back at the journey, I remember seeing the curiosity in front of science, the excitement at the initial light of breakthrough, the bitterness of working out the truth that seemed would never arrive. The fact that I arrived here today puts me in a great debt to many people. It is precisely them who appreciated my curiosity, echoed my excitement, and most importantly, diverted my desperate bitterness into persistence. I would like to take this opportunity truly express my gratitude to all of them.

First and foremost, my appreciation goes to my supervisor Wouter. I had the honor meeting Wouter when I attended an internship at Erusmas MC, DURING master program. And I was definitely thrilled when he accepted my application to pursue further study through a doctoral program. In the following years, his vast knowledge helped me focusing energy on the right path, his ever coming ideas kept inspiring my thinking, and his energetic working style stimulates everyone in the group. Beyond academic work, Wouter has a great patience and understanding while interacting with all group members including me and those kind support has been proven vital for an inexperienced scholar working on a 5 year pilot project and I found it directly responsible for my accomplishments today. Thank you for all the help in my academic and personal development.

I also want to thank Marc and Johan for being my evaluation committee member for my PhD project. Thank you for spending your time providing all those invaluable vision and suggestion at an early stage.

Modern Biomedical research could hardly be conducted without enormous lab work and data analysis. Every result in this thesis was obtained with the help and support from lab-mates and collaborators.

Thanks to Pernette and Diewertje for providing us with the plasmids and contributing to the lacO project. Thanks to Skoultchi and colleagues in his group for collaborating the H1 project and finalizing it into a paper. Thanks to Adriaan, Martin and Tison for helping me with the live imaging project.

Patrick, you are a mentor and friend to me. I feel very fortunate to have you in the lacO project. Through our daily/weekly discussion, I have learned so much knowledge/skills/thinking methodology from you, especially your impressive writing skills. It would be much tougher to complete my study without your help. Peter, I consider your appearance in this group was a big turning point in my project, a good one. Thanks for all the help with the experiments. Geert, Thanks for all the critical data analysis with your bioinformatics expertise. It is precisely your work that enabled me seeing the true meaning through the massive data. Thanks for being my paranimf. Elzo, it was great and challenging to work with you. Thanks for all the valuable advice and help with the Nature paper.

Erik, I feel very lucky to have you as my supervisor during my internship. Although it was a short internship, your vast knowledge and professional vision helped me developing interest in stem cell research. It was also fun to work with you in the following years. You were always the person for me to get the right answer☺. Petra, my "Lab-mates". Thanks for being my paranimf. We worked together all the way from Rotterdam to Utrecht. Thanks for all the help not only at work, but also extends to our shared interests and memories in life.

Sjoerd, it was great to have you around as a colleague and friend all those years. Thanks for sharing with me when I was stressed, depressed, frustrated. We also had a lot of fun together, thanks for encouraging me to learn Dutch and get know more about Dutch culture.

Paula, it was a joyful journey to both work and play with you, especially the last few months while concluding the projects and preparing thesis. It was a great comforting to have someone sharing all the feelings through this tough period. Yuva, Britta, Carlo and Erica, thanks for all the support and help. You make the office/lab life colorful and best wishes with finishing your projects.

Marjon, it was great to work with you and have you as my office neighbor. Thanks for all the help with the experiments. Mark and Hans, Thanks for all the help with cell culture, cloning and sorting.

&

To the new group postdocs members, Adrien, Annette and Christian, thanks for all the tips and advice.

To former group members, Blaise, Daan, Marieke, Robert-Jan and Harmen, thanks for their guidance and contribution at the beginning of my PhD stage.

Hubrecht is a great place to work. Tamara, it was fantastic to have you around. I started missing all the laugh we had after your relocation to Switzerland. Paul, Wensi, Flore, Dan, Leon, I was never bored with you guys. I also want to thank Stienenke and Stefan for their help with cell culture and sorting. Thank Yvonne and Rico from HR department for helping me with arranging resident permit.

If we say life without a mission is a journey without destination, then nobody likes a solo journey. It is not easy to start a new life far away from home. Therefore, I thank my friends Tian, Zhuoyi, Haowen, Chao, Jie, Sha, Jinwei and Rui, who not only make my life full of joy, but also make me brave.

I want to thank my parents. I consider myself lucky in a culture with relatively strong and stringent family ties. Unlike many other parents, they never impose their own dream or certain vision on me. On the contrary, they prepare my mentality in a healthy status, centering around kind and grateful to others, and encourage me to pursue my own dream. When I, their only child, decided to travel 9000 kilometers to study in Netherlands, all they check before agreeing was: is it a safe country to live for a girl? Their mere existence is an incomparable comfort to me and their love and understand put me in a debt I can never repay (谢谢，爸爸妈妈). Finally, to my husband Jiansong, thanks for truly understanding me and fully support. It was coincidence for me to live in Utrecht, but I believe it was a truly blessing fate that we met each other, got married, and had our lovely baby boy Feige. Thank you and I love you.

CURRICULUM VITAE

Yun Zhu was born on 4th November 1983 in **Nanyang**, China. After finishing her high school in home city in 2002, she was admitted to bachelor program of Shanghai JiaoTong University (**SJTU**) in Shanghai to study Plant Science and Technology. In 2006, after gaining her bachelor degree from SJTU, she came to Netherlands to for the master program of Molecular Cellular Biology in **Leiden University** in Leiden. Following the graduation from Leiden in 2008, equipped with a growing interest on genome study, she decided to pursue her academic career in **Hubrecht Institute** in Utrecht with the group of Prof. Wouter de Laat on Biomedical genomics. The results of her study are described in this thesis.

LIST OF PUBLICATIONS

Characterization and dynamics of pericentromere-associated domains in mice.
Wijchers PJ, Geeven G, Eyres M, Bergsma AJ, Janssen M, Verstegen M, **Zhu Y**, Schell Y, Vermeulen C, de Wit E, de Laat W.
Genome Res. 2015 Apr 16. pii: gr.186643.114.

The pluripotent genome in three dimensions is shaped around pluripotency factors.
de Wit E, Bouwman BA, **Zhu Y**, Klous P, Splinter E, Verstegen MJ, Krijger PH, Festuccia N, Nora EP, Welling M, Heard E, Geijsen N, Poot RA, Chambers I, de Laat W.
Nature. 2013 Sep 12;501(7466):227-31. doi: 10.1038/nature12420.

The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA.
Splinter E, de Wit E, Nora EP, Klous P, van de Werken HJ, **Zhu Y**, Kaaij LJ, van Ijcken W, Gribnau J, Heard E, de Laat W.
Genes Dev. 2011 Jul 1;25(13):1371-83. doi: 10.1101/gad.633311.

Local compartment changes and regulatory landscape alterations in histone H1 depleted cells
Geert Geeven[#], **Yun Zhu**[#], Byung Ju Kim[#], Boris A. Bartholdy[#], Seung-Min Yang, Todd S. Macfarlan, Wesley D. Gifford, Samuel L. Pfaff, Marjon J. A. M. Verstegen, Patrick J. Wijchers, John A. Stamatoyannopoulos, Arthur I. Skoultchi, & Wouter de Laat.
[#] equal contribution
Under review

Transcription-independent control of higher-order chromosome topology by trans-acting factors
Patrick J.Wijchers#, Peter H. L. Krijger#, **Yun Zhu**#, Diewertje G. E. Piebes, Pernette J. Verschure, Marjon J. A. M. Verstegen, Mark Janssen, Hans Teunissen, Geert Geeven & Wouter de laat.
[#] equal contribution
Manuscript in preparation