Chromatin conformation signatures: ideal human disease biomarkers?

Human health is related to information stored in our genetic code, which is highly variable even amongst healthy individuals. Gene expression is orchestrated by numerous control elements that may be located anywhere in the genome, and can regulate distal genes by physically interacting with them. These DNA contacts can be mapped with the chromosome conformation capture and related technologies. Several studies now demonstrate that gene expression patterns are associated with specific chromatin structures, and may therefore correlate with chromatin conformation signatures. Here, we present an overview of genome organization and its relationship with gene expression. We also summarize how chromatin conformation signatures can be identified and discuss why they might represent ideal biomarkers of human disease in such genetically diverse populations.

KEYWORDS: 3C-carbon copy ■ biomarker ■ chromatin conformation signature ■ chromatin structure ■ chromosome conformation capture ■ long-range regulation ■ transcription

Jennifer L Crutchley¹,
Xue Qing David Wang¹,
Maria A Ferraiuolo¹
& Josée Dostie^{1*}

¹Department of Biochemistry,
McGill University, 3655 Promenade
Sir-William-Osler, Room 814,
Montréal, Québec, Canada
*Author for correspondence:
Tel.: +1 514 398 4975
Fax: +1 514 398 7384
josee.dostie@mcgill.ca

Recent advances in DNA sequencing technology uncovered a tremendous diversity in the human genetic code. Indeed, several million nucleotides were found to differ between individuals, even in the healthy population [1,2]. What might be the impact of such variability on human health? Might sequence variations impart disease susceptibility or differential drug response? As human health is related to information stored in our genetic code, this enormous variability will have a significant impact by affecting the expression of genes. Sequence variation may target genes and regulatory DNA elements directly, or alter gene expression by affecting spatial genome organization. Although still poorly understood, spatial chromatin organization is emerging as an important mechanism to regulate the expression of genes. Therefore, understanding genome organization will be crucial to the development of optimal molecular targeted personalized therapies. In this article, we report on the relationship between gene expression and chromatin structure, beginning with a summary of human genome organization below.

Spatial genome organization *in vivo*

The ability to store, retrieve and translate instructions from the genetic code is essential to maintain life in all cells. This process is not trivial by any means in human cells given the size of our genome. In fact, understanding this process is not trivial even for much smaller genomes. The human genetic code is

composed of over 3 billion nucleotides, which when pieced together, would measure almost a meter in length. Therefore, our genome must be tightly packaged and organized in order to fit within each micron-sized nuclei. Packaging of the human genome is functional rather than random, and there are three defined hierarchical levels of organization (FIGURE 1) [3–5]. The first level of genome organization is characterized by the linear arrangement of genes and regulatory sequences (or ‘DNA elements’) along chromosomes. This first dimension includes clusters of genes and their regulatory DNA elements. Gene clusters composed of evolutionarily duplicated genes tend to encode proteins with similar functions and with tissue-specific expression patterns defined by their regulatory elements. Examples of this level of organization include the Hox gene clusters and α/β -globin loci, both of which will be further described in sections later.

The second level of genome organization is defined by the interaction between DNA and proteins. This second dimension is dominated by the relationship between genomic DNA and histones, where DNA is wrapped around nucleosomes to form the 10 nm chromatin fiber. At this level, chromatin appears as beads on a string, with beads corresponding to nucleosomes composed of two copies each of histone H2A, H2B, H3 and H4. Histones can be extensively modified post-translationally by acetylation, methylation, phosphorylation, sumoylation,

Figure 1. *In vivo* spatial genome organization. Three hierarchical levels of genome organization are illustrated from top to bottom. A gene cluster represents the first level of genome organization where the double helix is shown as yellow and blue strands. Transcriptional start sites indicated by arrows are highlighted with yellow circles. Highlighted in red is a nearby enhancer element regulating downstream genes. The 10-nm fiber is shown as an example of second level genome organization. A double helix wrapped around histone octamers represent nucleosomes. Further coiling of nucleosomes forms the 30-nm fiber. The third level of genome organization is represented as progressively larger and more compact chromatin fibers. Chromosome territories are highlighted in orange, green, violet, red, yellow and blue clusters. Intra- and inter-chromosomal contacts mediated by proteins are represented by red spheres. CT: Chromosome territory.

ADP-ribosylation and ubiquitinylation [6,7]. These epigenetic marks are mostly added to histone amino-terminal tails, and regulate their affinity to DNA and the recruitment of regulatory chromatin binding proteins. Histone modifications can also affect formation of the 30 nm chromatin fiber, which consists of a folded basic 10 nm fiber with nucleosomes stacked on top of each other.

Very little is known about genome organization beyond the 30 nm fiber, of which the *in vivo* structure remains to be established [8]. Even at this level of packaging, a stretched out 30 nm chromatin fiber with the DNA content

of an average chromosome would not fit into a nucleus. Therefore, additional folding and organization is essential for genome function. The third level of genome organization is defined by the packaging and spatial arrangement of chromatin in the nuclear space. This 3D organization is controlled by specialized proteins that bind and fold the 30-nm fiber into higher levels of organization such as loops. In addition to facilitating the accurate retrieval and translation of instructions from our genetic code, the spatial chromatin architecture of our genome is also used as a mechanism to regulate gene expression [9–11]. Indeed, it was shown

that DNA elements can regulate the expression of distal target genes by physically interacting with them [10]. This relatively recent discovery explains why the functional organization of the genome is not strictly linear along chromosomes and how DNA elements can regulate genes located very far away on the same or even on different chromosomes. Thus, long-range chromatin contacts in *cis* (intrachromosomal) or in *trans* (interchromosomal) can regulate gene expression by bringing regulatory elements in close physical proximity to target genes. Here, we refer to 'long-range' chromatin contacts from an empirical standpoint as interactions stronger than those originating from random collisions surrounding regions of interest. Long-range chromatin contacts were found to regulate genes from diverse cellular pathways, indicating that this form of control is a general regulation mechanism [12–24]. However, at least for some genomic regions, regulation through long-range DNA contacts has remained unclear [25–28]. Nonetheless, coregulated genes located far from each other or on different chromosomes also can co-localize and form foci in the nuclear space [29,30]. This type of organization likely participates in coordinating the proper timing and/or relative expression levels of various genes. 3D genome organization also includes positioning chromosomes into distinct territories within the nucleus, with gene-rich chromosomes at the center and gene-poor chromosomes near the periphery [3,31,32].

Measuring spatial genome organization

It had long been suspected that genes could be controlled over large genomic distances through physical contacts with DNA elements. However, this type of regulation mechanism has only been firmly demonstrated recently as a result of the development of various powerful technologies. Today, genome architecture can be studied with several approaches, which are usually combined for discovery (FIGURE 2 & TABLE 1). These techniques vary in resolution, throughput and cost, and together offer an unprecedented high-resolution view of our genome *in vivo*. Important information about spatial genome architecture has already come to light from these methodologies and includes the identification of intra- and interchromosomal contacts with roles in transcription or imprinting, the establishment of physical networks of coregulated genes, and an overall assessment of our genome architecture, including the existence of chromosome

territories. Implementation of these technologies will help compile a complete list of functional chromatin contacts, which may be impaired in human disease. Importantly, these techniques will identify unique and conserved structural features amongst the highly variable genomes of healthy individuals, which may impart differential disease susceptibility and response to drug treatment.

■ DNA-FISH

Until recently, DNA fluorescence *in situ* hybridization (DNA-FISH) was the main tool to measure chromatin contacts and other genomic features such as structural variations [33]. This technique is based on homologous sequence hybridization between an artificial DNA probe and the genomic DNA of cells chemically fixed on glass slides. The artificial DNA probe contains an epitope that can be specifically recognized by fluorescently labeled antibodies. Thus, hybridization sites can be visualized by epifluorescence microscopy and the position of multiple genomic regions can be measured simultaneously when using different combinations of probes and fluorescence tags.

As with any other approach, DNA-FISH offers both advantages and disadvantages to study genome organization. An important advantage of DNA-FISH is that it measures chromatin contacts in single cells. However, the resolution it provides is relatively low compared with newer sequencing-based methods, partly owing to the microscope's detection limits (see later). Although not the focus of this article, it is also important to note that advancements in fluorescence microscopy can now partly overcome these limitations by allowing resolutions beyond the Abbe limit [34]. Visualization of DNA-FISH fluorescence signals in chemically fixed cells requires that superimposed signals be deconvoluted in order to accurately estimate physical distances along the z-axis. Chemical fixation is known to alter cell morphology, which can introduce error in distance measurements. This drawback of DNA-FISH is alleviated by measuring physical distances between fluorescent probes in a large number of individual cells. Nevertheless, DNA-FISH remains the most important method to validate long-range functional chromatin contacts *in vivo*. This technique is well suited to study the dynamics of a few genomic regions or overall genome architecture and represents a perfect complement to the recently developed chromosome conformation capture (3C) and 3C-related technologies (see

Figure 2. Mapping spatial genome organization with chromosome conformation capture and related technologies. Five techniques used to map physical chromatin contacts at high resolution *in vivo* are illustrated from top to bottom. Chemical crosslinking of cells is a common first step to all approaches and is used to capture chromatin structure. Interacting DNA segments crosslinked by proteins are shown as yellow and green lines, and green and orange spheres, respectively. Fixed chromatin is either digested with a restriction enzyme (shown as scissors) or sheared by sonication (represented by concentric circles) to release crosslinked DNA fragments. Yellow and green arrows represent chromosome conformation capture (3C) primers. 5C primers used during the ligation-mediated amplification step are illustrated by yellow/black and green/gray lines, where black and gray moieties represent universal primer sequences. Y-shaped molecules represent antibodies. Biotinylated nucleotides are shown as red dots. Streptavidin beads are shown in purple. 3C: Chromosome conformation capture; 6C: Combined chromosome conformation capture ChIP cloning; ChIA-PET: Chromatin interaction analysis using paired-end tags; ChIP: Chromatin immunoprecipitation; IP: Immunoprecipitation; LMA: Ligation-mediated amplification.

later). Together, these approaches will likely lead to a better understanding of genome organization and function.

■ RNA-TRAP

RNA tagging and recovery of associated proteins (RNA-TRAP) is a technique reminiscent

of DNA-FISH introduced by Carter *et al.* in 2002 [35,36]. This approach captures the local DNA environment of an actively transcribed gene by combining aspects of FISH and chromatin immunoprecipitation (ChIP). Like DNA-FISH, RNA-TRAP uses homologous sequence hybridization of an artificial probe. However, instead of hybridizing on complementary genomic DNA, the probe binds to nascent unprocessed RNA at actively transcribed genes. During RNA-TRAP, the DNA probe is labeled with an epitope recognized by an antibody coupled to the horseradish peroxidase enzyme, which attaches biotin tags onto chromatin proteins in the immediate vicinity of nascent RNA transcripts. Thus, active transcription sites can be visualized with fluorescently labeled streptavidin. Additionally, active chromatin and associated proteins can be specifically purified by affinity chromatography and analyzed by quantitative PCR.

Unlike DNA-FISH, which provides a low resolution 'bird's eye' view of targeted chromatin components, RNA-TRAP can uncover in-depth information about the genomic environment of transcribed genes. However, the enzymatic step that tags proteins surrounding transcribed genes can trap proteins within a very large radius of activity. As such, RNA-TRAP captures the entire local environment of a given gene of interest rather than detecting direct physical interactions.

■ 3C

Chromosome conformation capture was initially developed to study the complete conformation of a chromosome in yeast [37]. 3C is now used as a standard research tool to analyze the organization of complex genomic domains and investigate the relationship between genome architecture and gene expression [38,39]. 3C can be divided into five experimental steps. The first step in conventional 3C is to chemically fix cells. This step captures interactions between DNA regions by crosslinking chromatin-bound histones and other associated proteins such as transcription factors. Thus, chemical fixation produces a snapshot of the 3D chromatin architecture *in vivo*. Chemical fixation is a common step in all techniques currently used to study genome organization. Although unavoidable, it is important to note that this step may still introduce artifacts that will be carried over in between approaches.

The second step of 3C consists of digesting the genomic DNA with enzymes. Enzymatic digestion of chemically fixed chromatin releases DNA fragments that were crosslinked as a result

of their physical proximity in the nuclear space. The third 3C step involves ligation of crosslinked DNA fragments. The DNA is ligated under conditions favoring intramolecular ligation of crosslinked fragments and minimizes random ligation. During the fourth step of 3C, the DNA is purified to remove all proteins and other contaminants. The resulting 3C library features pair-wise ligation products between DNA segments that were close to each other in the nuclear space regardless of their linear distance along the genome. The relative abundance of these ligation products is inversely proportional to the original 3D distance separating DNA segments and can therefore be used to reconstruct the spatial organization of the genome *in vivo*. The final 3C step consists of measuring the relative abundance of individual ligation products in the library. 3C library products are usually quantified by PCR amplification of ligation junctions and agarose gel detection. Alternatively, ligation junctions can be measured by TaqMan quantitative PCR

Table 1. Technologies employed to study spatial genome organization.

Technique	Genomic resolution/scale/throughput	Ref.
3C	High resolution Small genomic domains Low throughput	[37–39,59]
4C	High resolution Genomic environment surrounding a given region Low throughput	[44–46,98]
5C	High resolution Genome scale High throughput	[47–49,99]
6C	High resolution Genome-wide contacts associated with a given protein Intermediate throughput	[52,53]
3C-Loop	High resolution Genome-wide contacts associated with a given protein Low throughput	[42,43]
Hi-C	High resolution (proportional to sequencing depth) Genome-wide High throughput	[57]
ChIA-PET	High resolution Genome-wide contacts associated with a given protein (proportional to sequencing depth) High throughput	[54–56,100]
DNA-FISH	Low resolution Genome-wide Low throughput	[33,101,102]
RNA-TRAP	Intermediate resolution Genomic environment surrounding a given gene Low throughput	[35,36]

3C: Chromosome conformation capture; 4C: Circular chromosome conformation capture/ chromosome conformation capture on ChIP/open-ended chromosome conformation capture; 5C: Chromosome conformation capture carbon copy; 6C: Combined chromosome conformation capture ChIP cloning; ChIA-PET: Chromatin interaction analysis using paired-end tags; FISH: Fluorescence in situ hybridization; TRAP: Tagging and recovery of associated proteins.

or by melting curve analysis [40,41]. A major caveat of 3C and 3C-based technologies is that it generates datasets from cell populations and therefore features averaged interaction frequencies derived from various cell cycle states. Thus, these technologies yield averaged structural models rather than true structures. Although these models can be noisy, they remain useful to identify changes between cell states.

■ 3C-Loop

An immediate extension of 3C is the 3C-Loop technique, also known as the ChIP-loop method [42,43]. Like 3C, this technique also involves fixing cells to capture a 'snapshot' of *in vivo* genome architecture. However, 3C-Loop includes an immunoprecipitation step for a specific protein prior to ligation of crosslinked DNA fragments. 3C-Loop libraries are therefore enriched in ligation products previously bound by a protein of interest. Although removing unbound DNA fragments can substantially decrease background signals, this method requires prior knowledge of the regions bound by the specific proteins since contacts are measured by PCR with specific primers. Nonetheless, an important strength of 3C-Loop is that it allows identification of target proteins contributing to local chromatin looping. However, this method is most currently used to validate possible *cis/trans* long-range interactions.

■ 4C

The 4C techniques (circular chromosome conformation capture, chromosome conformation capture on chip, open-ended chromosome conformation capture) were developed to identify physical interactions genome-wide from any given genomic location [43–46]. Similar to 3C, 4C involves chemical fixation of cells, digestion of crosslinked DNA and ligation of crosslinked fragments to generate a library of DNA contacts. However, during 4C, genomic DNA is digested into very short fragments, which are then ligated under conditions promoting circularization of crosslinked DNA. 4C libraries consisting of short circularized ligation products are then purified as usual and amplified by reverse-PCR with primers nested at a specific genomic location. Reverse-PCR of 4C libraries thereby specifically amplifies all genomic regions physically interacting with the region of interest. Amplified DNA contacts are then identified on microarrays or by high-throughput DNA sequencing. Although DNA contacts from a fixed genomic location can be identified *ab initio* genome-wide at high resolution with 4C, this method is not

as quantitative as conventional 3C or 5C and should therefore be used mainly to identify interactions rather than quantify them.

■ 5C

The chromosome conformation capture carbon copy (5C) technology is also derived from 3C but allows quantitative simultaneous genome-wide detection of thousands of DNA contacts [47–51]. During 5C, a 3C library is first generated using the standard 3C protocol. However, instead of quantifying DNA contacts individually by PCR amplification and agarose gel detection, 3C libraries are first converted into 5C libraries and then analyzed on custom microarrays or by high-throughput DNA sequencing. 3C to 5C library conversion is achieved by a ligation-mediated amplification step involving annealing and ligation of primers corresponding to 3C ligation junctions. This ligation-mediated amplification step quantitatively detects 3C products specifically thereby creating a 'carbon copy' of DNA contacts, which is amplified by PCR and analyzed on microarrays or high-throughput DNA sequencing. Although 5C is very quantitative and somewhat high throughput, this approach does not identify DNA contacts without prior knowledge of regions involved since it relies on the ligation of primers at predicted 3C junctions.

■ 6C

The combined chromosome conformation capture ChIP cloning (6C) technique is also derived from 3C and is an immediate extension of the 3C-Loop approach [52,53]. 6C was developed to identify *cis* or *trans* long-range DNA interactions mediated by specific proteins without prior knowledge of the regions involved. As such, the 6C protocol is identical to 3C-Loop until the library purification step, but then uses a different approach to analyze libraries. During 6C, ligation products are first cloned into vectors rather than analyzed individually by PCR. Individual clones are then amplified and characterized by restriction digest analysis to identify those containing more than one DNA fragment. Clones with two or more fragments are then sequenced from both ends of the cloning vector to identify interacting sequences. Although 6C does not quantify DNA contacts like 3C-Loop, the combined cloning/sequencing shotgun approach qualitatively identifies long-range DNA interactions mediated by specific proteins.

The development of 3C technology by Dekker *et al.* in 2002 prompted the aggressive expansion of alternative 3C-derived approaches to study

high-resolution genome organization *in vivo*. These methods share similar protocols each with advantages and limitations but none are altogether genome-wide, quantitative, high throughput and applicable for *ab initio* contact identification. However, two state-of-the-art technologies developed over the past year fulfill these criteria. These techniques are called chromatin interaction analysis with paired-end tags (ChIA-PET) and Hi-C. These techniques are described later [54–57].

■ ChIA-PET

ChIA-PET was developed by the Ruan laboratory at the Genome Institute in Singapore [54,55]. This high-throughput technology represents a significant improvement upon the 3C-Loop and 6C technologies, as it quantitatively identifies chromatin contacts mediated by specific proteins across entire genomes simultaneously. ChIA-PET was first used to map the chromatin interaction network of estrogen receptor α (ER- α) in a breast cancer cell line. ChIA-PET combines two techniques: chromatin interaction analysis and high-throughput paired-end tag sequencing [56,58]. As with 3C, cells are first chemically fixed to capture *in vivo* chromatin contacts (FIGURE 2). However, instead of digesting genomic DNA with a restriction enzyme, the chromatin is sheared into small fragments by sonication. The fragmented chromatin is then immunoprecipitated with an antibody against any protein of interest. The DNA fragment ends are then repaired and ligated to epitope-labeled DNA linkers containing restriction sites. These products are further ligated under conditions favoring intermolecular ligation of crosslinked DNA, isolated by affinity with the epitope tag of linkers, and digested a second time into very short ChIA-PET fragments. These fragments are finally ligated to sequencing linkers, and the resulting ChIA-PET libraries are sequenced to map all intra- and interchromosomal contacts mediated by a given protein in the genome.

■ Hi-C

Whereas ChIA-PET was designed to identify the complete interactome of given proteins, Hi-C was developed jointly by the Dekker and Lander laboratories to measure all long-range genome-wide DNA contacts simultaneously [57]. As with any other 3C-derived approach, the first steps of Hi-C involve fixing cells and digesting the crosslinked chromatin with a restriction enzyme. However, instead of immediately ligating the DNA, the overhangs produced by the enzyme are first filled with nucleotides including one

that is labeled with an epitope tag. These products are then ligated under conditions favoring intermolecular ligation of crosslinked DNA fragments and sonicated to reduce the size of Hi-C products. These Hi-C products are next isolated by affinity chromatography through the epitope tags marking the junctions, and ligated to sequencing linkers. The Hi-C library is finally analyzed by high-throughput DNA sequencing, and all intra- and interchromosomal contacts mapped to the reference genome.

Chromatin conformation signatures

The development of high-resolution techniques to study spatial chromatin organization offers an unprecedented view of our genome *in vivo*. Implementation of the more recently developed genome-wide technologies such as ChIA-PET and Hi-C will eventually yield complete reconstructions of highly variable human genome architectures *in vivo*. Regardless, the earlier 3C-related techniques have already uncovered new structure-based mechanisms of gene regulation. These control mechanisms involve different types of physical contacts such as promoter–enhancer or insulator–enhancer interactions. Irrespective of the kind of contacts involved, the expression state of genes regulated by this type of control mechanism can be identified by its spatial chromatin organization. Furthermore, general gene expression patterns may also be associated with specific chromatin structures. We term the collection of DNA contacts associated with specific gene expression profiles chromatin conformation signatures (CCSs). We classify CCSs into four distinct categories to simplify discussion (FIGURE 3).

■ Local chromatin organization

Local chromatin organization refers to the chromatin compaction state of specific genomic locations at distances below looping detection range. This type of signature may significantly vary around promoter and enhancer sequences depending on their activity, and is a measure of the levels of random collisions surrounding elements. For example, active elements can be more open and yield less DNA contacts while silent ones can appear more compact and produce stronger interactions. This ‘opening’ of the chromatin was previously observed at the locus control region (LCR) in the β -globin locus [47]. There are instances, however, where local chromatin organization may remain constant regardless of cell type or cellular conditions. Such is the case of gene deserts devoid of transcription

Figure 3. Chromatin conformation signatures. Collections of DNA contacts representing chromatin conformation signatures are separated into four distinct categories for discussion. **(A)** Local chromatin organization refers to compaction levels at genomic sites. Organization of the 10-nm fiber at a transcribed gene is shown as described in **FIGURE 1**. Transcription machinery is represented by orange, green and yellow spheres. **(B)** Long-range intrachromosomal (*cis*) contacts. **(C)** Interchromosomal (*trans*) contacts. **(D)** Genomic environment. DNA from three chromosomes converges in the nuclear space to share regulatory factors and/or control elements and form contacts. A transcription factory is highlighted in yellow. The double helices from different chromosomes are shown as blue/yellow, red/yellow or purple/yellow strands. Rod-like folding of the double helix represents the 30-nm fiber. Dashed red arrows represent active transcription and blue spheres indicate enhancer-binding proteins.

activity, which are actually used as reference CCSs for sample comparison in conformation studies [59].

■ Intrachromosomal contacts

The functional organization of genes and DNA elements is not strictly linear along chromosomes. While transcription factor binding sites tend to localize at promoters, a given element may regulate distant genes without affecting the ones adjacent to it. Also, the relationship between genes and regulatory DNA elements is complex

in that each gene may be controlled by multiple elements and each element may control more than one gene. Furthermore, multiple DNA elements might function simultaneously or independently, dependent upon cellular conditions. Studies using 3C or 3C-related technologies have now shown that DNA elements can control distal genes by forming long-range intrachromosomal physical contacts with them. This type of signature was first identified in the β -globin locus between the LCR and actively transcribed globin genes [13]. Intrachromosomal contacts were found

to be essential and mediated by hematopoietic transcription factors. Although the β -globin locus was the first region shown to be regulated by this type of mechanism, functional intrachromosomal contacts have since been found to play important roles in the regulation of numerous other genes [12,13,15–24,55,60].

■ Interchromosomal contacts

In addition to intrachromosomal contacts, there are instances when genomic loci on separate chromosomes interact with each other in the nuclear space to control gene expression. The functional significance of these interchromosomal contacts is somewhat debated. Whereas functionally significant long-range contacts may occur between regulatory elements and target genes, other long-range interactions may simply be a consequence of genome compaction and bear no immediately apparent functional significance. Nonetheless, the first functional example of this type of CCS was identified between the promoter region of the IFN- γ gene on chromosome 10 and the regulatory regions of the Th2 cytokine locus on chromosome 11 [61]. This contact was confirmed by DNA-FISH and is thought to maintain both loci in an active state and allow for a rapid response upon T-cell activation to differentiate into Th1 and Th2 cell lineages by expression of either gene loci. Interchromosomal contacts have since been found between other genomic regions [14,27,55,62].

■ Genomic environment

Genomic environment refers to the concentration and composition of DNA sequences in the nuclear space surrounding a given genomic position. It was shown that coregulated genes often cluster together to share similar transcription factories irrespective of their linear genomic positions [30,55,63,64]. Altogether, these studies indicate that genomes are likely organized into dynamic networks of physical contacts bringing genes and regulatory elements in to close proximity to orchestrate gene expression. This model is further supported by a recent Hi-C study confirming the spatial proximity of small, gene-rich chromosomes [57]. Hi-C analysis has also generated unbiased long-range interaction maps of the human genome. These maps confirmed the existence of distinct chromosome territories and revealed that the genome is further divided into two types of spatial compartments, or genomic environments. The first environment contains active chromatin, which is typically gene rich and structurally open. The second environment

contains inactive gene-poor segments of chromatin that are structurally closed. Thus, the human genome appears to be generally organized into knot-free ‘fractal globule’ conformations (or environments) maximizing dense packaging and the ability to remain structurally dynamic.

CCSs as markers of gene expression states

Over the last few years it has become apparent that patterns of gene expression can be associated with distinct chromatin structures. These CCSs can be summarized by collections of DNA contacts, which may be complex and tissue specific. The study of gene clusters has been instrumental in deciphering this information. Gene clusters are ideal models for genome organization studies since they tend to encode highly regulated tissue-specific genes, which are sometimes controlled during development. Thus, these genomic regions can facilitate the identification of both transcription-dependent and tissue-specific CCSs.

■ The β -globin locus

The first, and by far best characterized gene cluster remains the β -globin locus. In humans, this locus contains a set of five developmentally regulated genes (*HBE*, *HBG2*, *HBG1*, *HBD* and *HBB*) that encode variants of the hemoglobin β chain (FIGURE 4A). These genes are almost exclusively expressed in erythrocytes and follow a very specific developmental pattern with *HBE* expressed during embryogenesis, *HBG1* and *HBG2* in the fetal phase, and *HBD* and *HBB* in adults [65]. The β -globin genes are controlled by an element, the LCR, which is situated approximately 25 kb upstream of the most proximal gene (*HBE*), and over 80 kb away from the farthest one (*HBB*). Although it was known for some time that the LCR is required to specifically activate each β -globin gene sequentially during development, the mechanism of this long-distance regulation remained unknown until 3C was applied to examine the cluster [13,66–68]. 3C analysis revealed that the LCR physically interacts specifically with actively transcribed genes and not with silent ones [69,70]. It was found that chromatin looping was mediated by erythroid-specific transcription factors, presumably by bridging enhancer sequences of the LCR to the promoters of transcribed genes. Indeed, looping between the LCR and adult β -globin genes was observed in definitive erythroid cells where the genes are expressed, but not in progenitor erythroid cells where they remain transcriptionally

Figure 4. Paradigm chromatin conformation signatures at transcriptionally regulated genomic loci. Gain of contacts does not equate to transcription activation. **(A)** Regulation of the β -globin locus in erythroid precursor cells (left) and definitive erythrocytes (right). The transcriptionally silent cluster (off) adopts a poised hub conformation maintained by proteins (red spheres). Adult β -globin gene transcription (on) is associated with contact formation between the locus control region (highlighted in red) and β -globin genes (yellow circles). **(B)** Regulation of the HoxA gene cluster. The transcriptionally silent locus features several chromatin loops (off). Transcription induction is associated with loss of contacts and unfolding of the cluster (on). Gray arrows marked with an X (red) indicate silent transcription start sites. A dashed red arrow represents active transcription. Transcription machinery is represented by orange, green and yellow spheres.

silent. Interestingly, the LCR was also found to bind and form a loop with sequences downstream of the cluster in both progenitor and definitive erythroid cells, but not in the brain where the cluster is always inactive and where no looping is ever detected. These results led to the active chromatin hub (ACH) model, whereby the β -globin cluster is suggested to adopt a basic conformation, which primes the cluster for activation specifically in erythroid cells. Thus, analysis of the β -globin cluster revealed the existence

of both transcription-dependent and tissue-specific CCSs, both of which can be important for transcription regulation.

■ The HoxA cluster

In the β -globin locus, the presence of contacts between the LCR and genes correlates with transcription activation. Similarly, clustering of the α - and β -globin genes in the nuclear space occurs when the genes are transcribed [30]. However, transcriptional activity does not necessarily

associate with the establishment of contacts. In fact, loss of contacts is sometimes correlated with transcription activation. A good example of this type of CCS was found in the *HoxA* cluster, which is also regulated in a tissue-specific manner during development (FIGURE 4B) [50].

The *Hox* genes are members of the evolutionary conserved homeobox superfamily that encode developmentally regulated transcription factors [71]. In humans, there are 39 *Hox* genes, which are organized into four clusters of 13 paralogue groups located on separate chromosomes. The *HoxA* cluster is located on chromosome 7 and encodes 11 transcription factors. During development, the spatial and temporal expression of these genes follows the order of their position along the chromosome. For example, *HoxA* genes located at the 3' end of the cluster are expressed more anteriorly in the embryo and earlier during development [72,73]. This colinearity and previous *in situ* hybridization studies strongly suggests that chromatin structure plays a central role in *Hox* regulation [74,75]. Perhaps even more interesting is the observation that transcriptional silencing is key to proper *Hox* function, since ectopic expression can lead to human disease. We have recently found that the *HoxA* cluster is organized into multiple discrete chromatin loops when transcriptionally silent and that DNA looping is absent when genes are actively transcribed [50,76]. Thus, gene clustering through chromatin looping appears to be a CCS of the transcriptionally silent *HoxA* gene cluster. Importantly, specific and sequential unfolding of these chromatin loops allowing access to the transcription machinery may hold the key to the developmental colinearity of *Hox* gene clusters.

Regulation of CCSs

While it is clear that DNA contacts are essential to regulate the expression of some genes, the factors forming and maintaining these contacts, and the pathways regulating them are very poorly understood. The recent identification of factors capable of forming chromatin loops genome-wide suggests that general CCS regulation pathways may also exist in addition to gene-specific control mechanisms. The CTCF and SATB1 proteins described below are two DNA looping factors known to integrate higher-order chromatin architecture with gene regulation.

■ CTCF

Strong evidence supports a role for the CCCTC-binding factor (CTCF) as a genome-wide CCS mediator and central regulator of spatial genome

organization [77,78]. First, CTCF is an insulator-binding protein shown to form loops at boundaries separating active and inactive chromatin domains, thereby maintaining different transcription activities in distinct nuclear compartments [79]. Second, CTCF was found to mediate functional chromatin loops within specific gene loci. For example, CTCF was shown to bind and bridge multiple regulatory elements in the mouse β -globin locus and mediate formation of the so-called ACH [68]. Indeed, conditional deletion of CTCF or disruption of regulatory elements with CTCF binding sites destabilized the ACH.

Probably the most characterized intrachromosomal CTCF contacts are the ones that regulate imprinting at the *Igf2/H19* locus [80–82]. The *H19* and *Igf2* genes are located approximately 100 kb away from each other on human chromosome 11. Between the genes and proximal to *H19*, a DNA sequence called imprinting control region (ICR) is a well-known CTCF-binding site. On the maternal allele, CTCF can bind to the ICR and forms multiple loops with regions along the locus that prevents *Igf2* from physically interacting with its enhancer sequence. By contrast, on the paternal allele, CTCF binding to the ICR is abolished by DNA methylation and the *Igf2* gene is able to form a long-range DNA contact with its enhancer. Thus, CTCF binding and loop formation is essential to regulate the entire locus and ensure that the *Igf2* gene is expressed only from the paternal allele.

In addition to mediating intrachromosomal contacts, CTCF can also form functional interchromosomal interactions between coregulated genomic regions. For example, the *Igf2/H19* and *Wsb1/Nf1* gene loci were shown to interact in a CTCF-dependent manner [14]. Indeed, CTCF depletion or deletion of the regulatory element required for CTCF binding abolished the interaction and altered gene expression at the *Wsb1/Nf1* locus. CTCF might therefore direct distant genomic segments to common transcription factories by mediating interchromosomal contacts.

■ SATB1

Special AT-rich sequence binding protein-1 (SATB1) is another relatively well-characterized master organizer of *in vivo* chromatin structure [83–85]. This protein binds to specific DNA sequences termed matrix attachment regions, and anchors the genome to the nuclear matrix through series of distinct loops. The

chromatin loops formed by SATB1 participate in establishing the overall genome architecture and are known to function in gene regulation. SATB1 regulates gene expression by at least two types of mechanisms. First, it is known to form distinct chromatin structures that selectively tether DNA elements to specific nuclear compartments. For example, SATB1 was shown to promote enhancer activity over long distances by forming cage-like DNA networks that move distal elements and target genes in close proximity. The SATB1 networks also appear to activate transcription by segregating heterochromatin to other nuclear compartments. Second, SATB1 can inhibit transcription by recruiting chromatin modifiers, such as histone deacetylases, and remodelers, such as ATP-dependent chromatin-assembly factor and imitation switch [86].

Special AT-rich sequence binding protein-1 is cell-type specific, and its role in regulating the Th2 cytokine locus is well documented in thymocytes. The mouse Th2 cytokine locus located on chromosome 11 measures approximately 200 kb, and encodes the IL-5, IL-4 and IL-13 genes. Expression of these genes is coordinated upon Th2 cell activation and it was shown that following activation, the locus folds into numerous small DNA loops anchored at the base by SATB1. RNAi knockdown experiments demonstrated that SATB1 does not simply organize the locus into distinct loops. Instead, SATB1 was also found to be required for the expression of the interleukin genes themselves and of the c-Maf transcription factor regulating the locus [87].

Genome variability, CCSs & human health

Variations in the human genetic code are very abundant. Although some variations may already be linked to human disease, the full impact of this diversity on human health is currently unknown. Human genome sequence variations can take many forms, ranging in size from single nucleotides to large chromosomal segments. Variants are classified into several distinct groups according to their size, and include single nucleotide polymorphisms (SNPs), insertion/deletions (indels), copy number variations (CNVs) and structural variations (SVs; insertions, deletion and inversions) (TABLE 2). Any type of sequence variation may change gene expression patterns and result in altered CCSs (FIGURE 5). Thus, variations of the human genome may bear distinct CCSs and identify specific gene expression profiles.

Single nucleotide polymorphisms are single nucleotide variations that may be substituted, deleted or inserted in the genome. Although SNPs are generally not pathogenic, a number have been associated with disease. They may be present within coding or noncoding gene sequences, or in intergenic regions. SNPs in gene coding regions may be synonymous and have no effect on protein function. Alternatively, SNPs can be nonsynonymous (nsSNPs) and produce protein variants. The effect of nsSNPs on protein function can range from none to the production of nonfunctional proteins. Likewise, intergenic SNPs may also bear multiple consequences, ranging from none to disrupting genome function. For example, SNPs at enhancer elements may have a serious impact on genome regulation (FIGURE 5B). SNPs may disrupt the binding sites of transcription factors or of structural proteins, alter transcription patterns and the formation of proper CCSs. Such is the case for the inherited rs6983267 SNP variant associated with colorectal cancer pathogenesis. This variant was shown to modify an enhancer sequence found to physically interact with the MYC proto-oncogene [88]. This intergenic SNP increased binding of the transcription factor 7-like 2 (TCF7L2) and enhanced MYC expression in colorectal cancer cells.

Structural variations (deletions, insertion and deletions) of large DNA segments could be particularly detrimental to proper genome function (FIGURE 5C, D & E). Like SNPs, structural variants can be found anywhere in the genome of healthy individuals. They can affect genes directly or alter regulatory DNA elements. For example, intergenic deletions may eliminate enhancer sequences and affect the expression pattern of one or multiple constitutive and/or regulated genes. Such SVs may induce local chromatin changes at gene promoters, alter intra- or interchromosomal contacts, and even modify the genomic environment of genes. Likewise, large intergenic insertions may change gene expression by disrupting regulatory elements or chromatin structures essential for transcription. Chromosomal inversions, either balanced or unbalanced, may also affect gene expression, particularly by altering genomic environments of genes. For example, a large balanced inversion might displace enhancer elements from shared transcription factories and affect regulation of multiple genes under specific cellular conditions. Thus, CCSs associated with even single SV may be very complex.

Copy number variations are of particular interest to disease pathogenesis because of their large size and effect on the overall genomic

environment. CNVs are classified as segments of DNA spanning 1 kb to several megabases in size, for which copy-number differences were observed between two or more genomes of the same species. These segments can be copy-number gains or losses of gene coding or intergenic genomic DNA. Intergenic CNV gain of DNA elements may directly modify gene expression and manifest complex CCSs of altered local chromatin structure and chromosomal contacts (FIGURE 5F). Alternatively, a CNV may contribute to the pathogenesis of a disease by altering the location of genes with respect to regulatory elements. For example, a CNV loss in close proximity to an enhancer might displace elements from their target genes and prevent formation of the correct CCS required for gene activation or silencing. An interesting example of this type of chromatin structure-induced altered transcription was found at the 4q35 locus in patients suffering from fascioscapulohumeral muscular dystrophy [22,89]. This dominant neuromuscular disorder is linked to the partial deletion of a polymorphic repeat region known as D4Z4 located in the subtelomeric region of chromosome 4q. Whereas the 3.3 kb D4Z4 repeat is present at up to 200 copies in healthy individuals, less than 10 are usually found in fascioscapulohumeral muscular dystrophy patients. Partial D4Z4 deletion was shown to prevent anchoring of a nearby matrix attachment region sequence to the nuclear matrix, which usually form distinct looped domains restricting 3D contacts between genes and enhancer sequences. Thus, neighboring genes are aberrantly overexpressed as a consequence of sequence variation-induced altered spatial chromatin organization. Chromatin architecture-induced aberrant transcription is not likely linked to just a few human disorders. With the recent influx of evidence indicating that multiple rare *de novo*, and inherited CNVs contribute to the genetic component of vulnerability to neuropsychiatric disorders such as autism spectrum disorder and schizophrenia [90–96], it will be interesting to see whether chromatin organization plays a role in these complex disorders.

In addition to changes in genome sequence, deregulation of factors regulating chromatin architecture may also play a significant role in disease pathogenesis. For example, the genome organizer SATB1 was found to be involved in breast cancer. Han and colleagues showed that reducing SATB1 protein levels by RNAi altered the expression of over 1000 genes and reversed the process of tumorigenesis [97]. Indeed, SATB1

Table 2. Human genome sequence variations.

Variation	Size	Ref.
Single nucleotide polymorphism	Single base pair	[103–106]
Insertion/deletions (indels)	Up to 1 kb	[107,108]
Copy number variation	1 kb or larger Mapped from 443 bp and larger copy number variations	[90]
Structural variations: Insertions Deletions Inversions	kb to Mb	[91,108,109]

knockdown was found to restore breast-like acinar polarity, and to inhibit tumor growth and metastasis. Furthermore, SATB1 was found to delineate specific chromatin modifications at target gene loci, which directly upregulated metastasis-associating genes and downregulated tumor-suppressor genes. SATB1 might therefore play a role in tumorigenesis by reprogramming chromatin organization and transcription profiles of cancer cells to promote growth and metastasis. Thus, regulation of chromatin structure and organization might contribute to disease pathogenesis and represent valuable biomarkers. The continual application of 3C-related technologies will likely uncover the mechanisms regulating CCS formation and reveal their impact on genome function.

CCSs as ‘ideal’ biomarkers?

The recent developments in gene sequencing, targeted therapies and molecular diagnostics are leading to a more personalized approach for the treatment of human disease. The identification of biomarkers with these technologies will continually improve personalized therapy as it will identify the threat of disease, if a disease is present, its severity and its response to drug treatment. For instance, biomarkers can provide the confirmation of a disease status necessary for correct treatment. In the case of chronic disease where individuals may require treatment for a long period of time, biomarkers may be critical for the timely identification and classification of a disease. Also, in the event of an early symptom-free phase, such as in Alzheimer’s disease, biomarkers may allow preventive treatment.

The current need for more accurate disease diagnosis and prognosis increases the priority to identify new biomarkers. Markers that can predict response or resistance to drug therapies are desirable and will help single out patients susceptible to severe adverse drug reactions and reduce the risk of treatment failure. The

Figure 5. Sequence variations can alter spatial genome organization and change gene expression. Any type of sequence variation can alter chromatin structure and either induce or inhibit transcription. A hypothetical enhancer/gene interchromosomal contact required for transcription is used as an example. Transcription inhibition is shown as the outcome of chromatin structure changes for all variants except CNVs. **(A)** DNA contact and gene expression of a healthy individual with a linear reference genome structure. **(B)** SNP can alter DNA binding site of a transcription factor and impair transcription. **(C)** DNA deletion can remove transcription factor binding sites. **(D & E)** DNA insertions or inversions can displace enhancer elements from their genomic environment. **(F)** CNVs can multiply the number of transcription factor binding sites and stimulate transcription. Enhancer-binding protein is shown as a blue sphere. Transcription machinery is illustrated by orange, green and yellow spheres. Red spheres are structural proteins added to indicate double helix orientation. Gray arrows marked with an X (red) indicate silent transcription start sites. A dashed red arrow represents active transcription. CNV: Copy number variation; SNP: Single nucleotide polymorphism.

variability of the human genome will likely hamper identification of the genetic components of diseases or render disease classification problematic. One of the advantages of using CCSs as biomarkers is that multiple linear variations may be integrated into unique spatial signatures. For example, combinations of SNPs, apparently nonpathogenic SVs and CNVs may together induce alternative genome conformations affecting an individual's ability to optimally regulate

gene expression and cope with environmental stresses. Thus, CCSs can potentially capture this 'genomic behavior' that may not be otherwise apparent in 'static' patient samples where cells were not subjected to the relevant environment conditions prior to analysis. Moreover, gene expression profiles such as steady-state mRNA levels or protein output can be uninformative when slight changes in multiple gene expression levels contribute to a given pathology. Small

gene expression changes could be discarded in routine transcriptome analysis but could still have important effects on cellular function. Alternative mRNA splicing patterns could also contribute to specific cell states. These changes would not be detected in regular mRNA profiling. Similarly, changes in the expression of regulatory noncoding RNAs such as miRNAs would not be detected in regular transcriptome analysis. Classifying components of gene regulation mechanisms as biomarkers is appropriate since they can be the underlying cause of human disease. As chromatin structure can regulate gene expression, CCSs may simply represent structural signatures of gene expression. Also, since CCSs are manifestations of gene expression profiles, they offer the advantage of identifying gene expression states regardless of the nature of the

mechanisms involved. Finally, CCSs uniquely identify 3D gene regulation mechanisms. Thus, aberrant gene expression resulting from changes in the positioning of genes in space may only be identified through CCSs.

Conclusion & future perspective

Linear genome structures vary significantly between individuals, even in the healthy population [92,93]. Since genomes appear organized into dynamic 3D networks of physical contacts, spatial chromatin organization is likely to be shaped by the linear arrangement of genes and regulatory DNA elements. Consequently, the nuclear environment of individual genes may be affected by linear sequence variations and lead to improper gene expression. For this reason, understanding how the human genome is

Executive summary

Overview

- Recent advances in genomics revealed tremendous human genome sequence diversity in the healthy population.
- Genome variability may impart disease complexity and differential drug response, or disease susceptibility.
- Sequence variation can alter gene expression either by targeting genes and regulatory DNA elements directly or by affecting genome organization.
- Understanding genome organization is crucial for optimal personalized therapies.

Summary of genome organization

- There are three hierarchical levels of genome organization:
 - The linear arrangement of genes and DNA elements along chromosomes
 - The association of genomic DNA with proteins and formation of chromatin
 - The packaging and organization of chromatin in the nuclear space
- Genome organization in the nuclear space is not random – it is functional.
- Spatial genome organization is an important mechanism to regulate gene expression.

Measuring spatial genome organization

- Genome architecture can be studied with several approaches.
- Available techniques vary in resolution, throughput, genomic coverage and cost.
- Current methodologies complement each other and should be combined for discovery.
- Chromosome conformation capture (3C)-related techniques offer an unprecedented high-resolution view of our genome.
- Specific DNA contacts can be measured with 3C-derived approaches.

Chromatin conformation signatures

- Chromatin conformation signatures (CCSs) are collections of DNA contacts associated with specific gene expression states.
- There are four types of CCSs:
 - Local chromatin organization
 - Intrachromosomal contacts
 - Interchromosomal contacts
 - Genomic environment
- CCSs can be complex and include several types of DNA contacts.

Genome variability, CCSs & human health

- Variations in the human genetic code are abundant and can be linked to disease.
- Gene expression patterns may be altered by sequence variations.
- Any type of variation may affect gene expression by altering genome organization.
- Sequence variations may bear distinct CCSs.
- CCSs help explain disease complexity and susceptibility, or differential drug response.

CCSs as ideal biomarkers

- CCSs may integrate multiple variations into single signatures.
- CCSs may identify gene expression states regardless of mechanisms.
- CCSs uniquely identify 3D mechanisms of gene regulation.

structured *in vivo* is key to understanding transcription regulation and other processes such as imprinting and DNA replication.

Chromosome conformation capture-related techniques offer an unprecedented view of genome organization at the ultrastructural level. The recent development of the ChIA-PET and Hi-C technologies has promoted chromosome conformation research to an entirely higher level. Constructions of long-range interaction maps of human genomes have already begun to emerge from these technologies. For the first time, a more 'top-down' approach to studying the role of genome architecture in the regulation of genes appears feasible. Currently, the main issue with the Hi-C and ChIA-PET technology is the high cost of high-throughput sequencing. As such, high-throughput spatial reconstruction of human genome libraries is not immediately

possible. However, as with any technology, the cost will decrease and sequencing will eventually be affordable enough for CCS screening in large cohorts. Nonetheless, mapping physical networks will unquestionably continue to uncover the relationship between CCSs and genome function, and provide a better understanding of human disease pathogenesis.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Bibliography

Papers of special note have been highlighted as:

■ of interest

■ of considerable interest

- Pinto D, Marshall C, Feuk L, Scherer SW: Copy-number variation in control population cohorts. *Hum. Mol. Genet.* 16(Spec No. 2), R168–R173 (2007).
- Lee C, Scherer SW: The clinical context of copy number variation in the human genome. *Expert Rev. Mol. Med.* 12, E8 (2010).
- Fraser P, Bickmore W: Nuclear organization of the genome and the potential for gene regulation. *Nature* 447(7143), 413–417 (2007).
- Babu MM, Janga SC, De Santiago I, Pombo A: Eukaryotic gene regulation in three dimensions and its impact on genome evolution. *Curr. Opin. Genet. Dev.* 18(6), 571–582 (2008).
- Cook PR: A model for all genomes: the role of transcription factories. *J. Mol. Biol.* 395(1), 1–10 (2010).
- Berger SL: The complex language of chromatin regulation during transcription. *Nature* 447(7143), 407–412 (2007).
- Kouzarides T: Chromatin modifications and their function. *Cell* 128(4), 693–705 (2007).
- Tremethick DJ: Higher-order structures of chromatin: the elusive 30 nm fiber. *Cell* 128(4), 651–654 (2007).
- Woodcock CL: Chromatin architecture. *Curr. Opin. Struct. Biol.* 16(2), 213–220 (2006).
- West AG, Fraser P: Remote control of gene transcription. *Hum. Mol. Genet.* 14(Spec No. 1), R101–R111 (2005).
- Gondor A, Ohlsson R: Chromosome crosstalk in three dimensions. *Nature* 461(7261), 212–217 (2009).
- Spilianakis CG, Flavell RA: Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* 5(10), 1017–1027 (2004).
- Tolhuis B, Palstra RJ, Splinter E, Grosveld F, De Laat W: Looping and interaction between hypersensitive sites in the active β -globin locus. *Mol. Cell* 10(6), 1453–1465 (2002).
- Ling JQ, Li T, Hu JF *et al.*: CTCF mediates interchromosomal colocalization between IGF2/H19 and WSB1/NF1. *Science* 312(5771), 269–272 (2006).
- Liu Z, Garrard WT: Long-range interactions between three transcriptional enhancers, active V κ gene promoters, and a 3' boundary sequence spanning 46 kilobases. *Mol. Cell Biol.* 25(8), 3220–3231 (2005).
- Murrell A, Heeson S, Reik W: Interaction between differentially methylated regions partitions the imprinted genes IGF2 and H19 into parent-specific chromatin loops. *Nat. Genet.* 36(8), 889–893 (2004).
- Lanzuolo C, Roure V, Dekker J, Bantignies F, Orlando V: Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 9(10), 1167–1174 (2007).
- Jiang H, Peterlin BM: Differential chromatin looping regulates CD4 expression in immature thymocytes. *Mol. Cell Biol.* 28(3), 907–912 (2008).
- Tsytzykova AV, Rajsbaum R, Falvo JV, Ligeiro F, Neely SR, Goldfeld AE: Activation-dependent intrachromosomal interactions formed by the *TNF* gene promoter and two distal enhancers. *Proc. Natl Acad. Sci. USA* 104(43), 16850–16855 (2007).
- Ju Z, Volpi SA, Hassan R *et al.*: Evidence for physical interaction between the immunoglobulin heavy chain variable region and the 3' regulatory region. *J. Biol. Chem.* 282(48), 35169–35178 (2007).
- D'haene B, Attanasio C, Beysen D *et al.*: Disease-causing 7.4 kb *cis*-regulatory deletion disrupting conserved non-coding sequences and their interaction with the foxl2 promoter: implications for mutation screening. *PLoS Genet.* 5(6), e1000522 (2009).
- Petrov A, Allinne J, Pirozhkova I, Laoudj D, Lipinski M, Vassetzky YS: A nuclear matrix attachment site in the 4q35 locus has an enhancer-blocking activity *in vivo*: implications for the facio-scapulo-humeral dystrophy. *Genome Res.* 18(1), 39–45 (2008).
- Dmitriev P, Lipinski M, Vassetzky YS: Pearls in the junk: dissecting the molecular pathogenesis of facioscapulohumeral muscular dystrophy. *Neuromuscul. Disord.* 19(1), 17–20 (2009).
- Sexton T, Bantignies F, Cavalli G: Genomic interactions: Chromatin loops and gene meeting points in transcriptional regulation. *Semin. Cell. Dev. Biol.* 20(7), 849–855 (2009).
- Lasalle JM, Lalande M: Homologous association of oppositely imprinted chromosomal domains. *Science* 272(5262), 725–728 (1996).
- Teller K, Solovei I, Buiting K, Horsthemke B, Cremer T: Maintenance of imprinting and nuclear architecture in cycling cells. *Proc. Natl Acad. Sci. USA* 104(38), 14970–14975 (2007).

- 27 Hu Q, Kwon YS, Nunez E *et al.*: Enhancing nuclear receptor-induced transcription requires nuclear motor and *LSD1*-dependent gene networking in interchromatin granules. *Proc. Natl Acad. Sci. USA* 105(49), 19199–19204 (2008).
- 28 Kocanova S, Kerr EA, Rafique S *et al.*: Activation of estrogen-responsive genes does not require their nuclear co-localization. *PLoS Genet.* 6(4), E1000922 (2010).
- 29 Osborne CS, Chakalova L, Brown KE *et al.*: Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36(10), 1065–1071 (2004).
- 30 Schoenfelder S, Sexton T, Chakalova L *et al.*: Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* 42(1), 53–61 (2010).
- **Describes the RNA-TRAP technique for the first time.**
- 31 Cremer T, Cremer C: Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* 2(4), 292–301 (2001).
- 32 Bolzer A, Kreth G, Solovei I *et al.*: Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol.* 3(5), E157 (2005).
- **Describes the 3C technique for the first time.**
- 33 Simonis M, De Laat W: FISH-eyed and genome-wide views on the spatial organisation of gene expression. *Biochim. Biophys. Acta.* 1783(11), 2052–2060 (2008).
- 34 Fernandez-Suarez M, Ting AY: Fluorescent probes for super-resolution imaging in living cells. *Nat. Rev. Mol. Cell. Biol.* 9(12), 929–943 (2008).
- 35 Carter D, Chakalova L, Osborne CS, Dai Y-F, Fraser P: Long-range chromatin regulatory interactions *in vivo*. *Nat. Genet.* 32(4), 623–626 (2002).
- 36 Bulger M, Groudine M: Trapping enhancer function. *Nat. Genet.* 32(4), 555–556 (2002).
- 37 Dekker J, Rippe K, Dekker M, Kleckner N: Capturing chromosome conformation. *Science* 295(5558), 1306–1311 (2002).
- 38 Miele A, Dekker J: Mapping *cis*- and *trans*-chromatin interaction networks using chromosome conformation capture (3C). *Methods Mol. Biol.* 464, 105–121 (2009).
- 39 Miele A, Gheldof N, Tabuchi TM, Dostie J, Dekker J: Mapping chromatin interactions by chromosome conformation capture (3C). In: *Current Protocols in Molecular Biology*. Ausubel FM, Brent R, Kingston RE *et al.* (Eds). John Wiley & Sons, Hoboken, NJ, USA (2006).
- **Along with references [40] and [41] describes a variation of the 4C approach for the first time.**
- 40 Hagege H, Klous P, Braem C *et al.*: Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat. Protoc.* 2(7), 1722–1733 (2007).
- **Along with references [39] and [41] describes a variation of the 4C approach for the first time.**
- 41 Abou El Hassan M, Bremner R: A rapid simple approach to quantify chromosome conformation capture. *Nucleic Acids Res.* 37(5), E35 (2009).
- **Along with references [39] and [40] describes a variation of the 4C approach for the first time.**
- 42 Horike S-I, Cai S, Miyano M, Cheng J-F, Kohwi-Shigematsu T: Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* 37(1), 31–40 (2005).
- **Describes the 5C technique for the first time.**
- 43 Simonis M, Kooren J, De Laat W: An evaluation of 3C-based methods to capture DNA interactions. *Nat. Meth.* 4(11), 895–901 (2007).
- 44 Zhao Z, Tavoosidana G, Sjolinder M *et al.*: Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* 38(11), 1341–1347 (2006).
- 45 Würtele H, Chartrand P: Genome-wide scanning of HOXB1-associated loci in mouse ES cells using an open-ended chromosome conformation capture methodology. *Chromosome Res.* 14(5), 477–495 (2006).
- 46 Simonis M, Klous P, Splinter E *et al.*: Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* 38(11), 1348–1354 (2006).
- 47 Dostie J, Richmond TA, Arnaout RA *et al.*: Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16(10), 1299–1309 (2006).
- 48 Dostie J, Zhan Y, Dekker J: High-throughput mapping of chromatin interactions using 5C technology. In: *Current Protocols in Molecular Biology*. Ausubel FM, Brent R, Kingston RR *et al.* (Eds). John Wiley & Sons, Hoboken, NJ, USA (2007).
- 49 Dostie J, Dekker J: Mapping networks of physical interactions between genomic elements using 5C technology. *Nat. Protoc.* 2(4), 988–1002 (2007).
- 50 Fraser J, Rousseau M, Shenker S *et al.*: Chromatin conformation signatures of cellular differentiation. *Genome Biol.* 10(4), R37 (2009).
- **First study using the ChIA-PET approach.**
- 51 Van Berkum NL, Dekker J: Determining spatial chromatin organization of large genomic regions using 5C technology. *Methods Mol. Biol.* 567, 189–213 (2009).
- 52 Tiwari VK, Cope L, Mcgarvey KM, Ohm JE, Baylin SB: A novel 6C assay uncovers polycomb-mediated higher order chromatin conformations. *Genome Res.* 18(7), 1171–1179 (2008).
- **Describes the Hi-C technique for the first time.**
- 53 Tiwari VK, Baylin SB: Combined 3C-chip-cloning (6C) assay: a tool to unravel protein-mediated genome architecture. *Cold Spring Harb. Protoc.* 2009(3), pdb. prot5168 (2009).
- 54 Fullwood MJ, Ruan Y: Chip-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* 107(1), 30–39 (2009).
- 55 Fullwood MJ, Liu MH, Pan YF *et al.*: An oestrogen-receptor-[AGR]-bound human chromatin interactome. *Nature* 462(7269), 58–64 (2009).
- 56 Fullwood MJ, Han Y, Wei CL, Ruan X, Ruan Y: Chromatin interaction analysis using paired-end tag sequencing. *Curr. Protoc. Mol. Biol.* 15, 21–25 (2010).
- **Identifies functional interchromosomal DNA contacts at high resolution for the first time.**
- 57 Lieberman-Aiden E, Van Berkum NL, Williams L *et al.*: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950), 289–293 (2009).
- 58 Fullwood MJ, Wei C-L, Liu ET, Ruan Y: Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* 19(4), 521–532 (2009).
- 59 Dekker J: The 3C's of chromosome conformation capture: controls, controls, controls. *Nat. Meth.* 3(1), 17–21 (2006).
- 60 Hakim O, John S, Ling JQ, Biddie SC, Hoffman AR, Hager GL: Glucocorticoid receptor activation of the CIZ1-LCN2 locus by long range interactions. *J. Biol. Chem.* 284(10), 6048–6052 (2009).
- 61 Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA: Interchromosomal associations between alternatively expressed loci. *Nature* 435(7042), 637–645 (2005).

- 62 Lomvardas S, Barnea G, Pisapia DJ, Mendelsohn M, Kirkland J, Axel R: Interchromosomal interactions and olfactory receptor choice. *Cell* 126(2), 403–413 (2006).
- 63 Sexton T, Schober H, Fraser P, Gasser SM: Gene regulation through nuclear organization. *Nat. Struct. Mol. Biol.* 14(11), 1049–1055 (2007).
- 64 Osborne CS, Chakalova L, Mitchell JA *et al.*: Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol.* 5(8), e192 (2007).
- 65 Trimborn T, Gribnau J, Grosveld F, Fraser P: Mechanisms of developmental control of transcription in the murine α - and β -globin loci. *Genes Dev.* 13(1), 112–124 (1999).
- 66 Palstra RJ, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, De Laat W: The β -globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* 35(2), 190–194 (2003).
- 67 Lewis EB: A gene complex controlling segmentation in drosophila. *Nature* 276(5688), 565–570 (1978).
- 68 Splinter E, Heath H, Kooren J *et al.*: CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev.* 20(17), 2349–2354 (2006).
- 69 De Laat W, Grosveld F: Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* 11(5), 447–459 (2003).
- 70 Vakoc C, Letting DL, Gheldof N *et al.*: Proximity among distant regulatory elements at the β -globin locus requires GATA-1 and FOG-1. *Mol. Cell* 17(3), 453–462 (2005).
- 71 Krumlauf R: Hox genes in vertebrate development. *Cell* 78(2), 191–201 (1994).
- 72 Kmita M, Duboule D: Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301(5631), 331–333 (2003).
- 73 Duboule D, Morata G: Colinearity and functional hierarchy among genes of the homeotic complexes. *Trends Genet.* 10(10), 358–364 (1994).
- 74 Morey C, DA Silva NR, Perry P, Bickmore WA: Nuclear reorganization and chromatin decondensation are conserved, but distinct, mechanisms linked to *HOX* gene activation. *Development* 134(5), 909–919 (2007).
- 75 Chambeyron S, Bickmore WA: Chromatin decondensation and nuclear reorganization of the *HOXB* locus upon induction of transcription. *Genes Dev.* 18(10), 1119–1130 (2004).
- 76 Ferraiuolo MA, Rousseau M, Miyamoto C *et al.*: The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* (2010) (In Press).
- 77 Filippova GN, Fagerlie S, Klenova EM *et al.*: An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. *Mol. Cell Biol.* 16(6), 2802–2813 (1996).
- 78 Phillips JE, Corces VG: CTCF: master weaver of the genome. *Cell* 137(7), 1194–1211 (2009).
- 79 Szabo PE, Tang SH, Silva FJ, Tsark WM, Mann JR: Role of CTCF binding sites in the IGF2/H19 imprinting control region. *Mol. Cell Biol.* 24(11), 4791–4800 (2004).
- 80 Kanduri C, Pant V, Loukinov D *et al.*: Functional association of CTCF with the insulator upstream of the H19 gene is parent of origin-specific and methylation-sensitive. *Curr. Biol.* 10(14), 853–856 (2000).
- 81 Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM: CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/IGF2 locus. *Nature* 405(6785), 486–489 (2000).
- 82 Bell AC, Felsenfeld G: Methylation of a CTCF-dependent boundary controls imprinted expression of the *IGF2* gene. *Nature* 405(6785), 482–485 (2000).
- 83 Galande S, Purbey PK, Notani D, Kumar PP: The third dimension of gene regulation: organization of dynamic chromatin loopscape by SATB1. *Curr. Opin. Genet. Dev.* 17(5), 408–414 (2007).
- 84 Cai S, Lee CC, Kohwi-Shigematsu T: SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* 38(11), 1278–1288 (2006).
- ■ **Demonstrates how a single nucleotide polymorphism can change chromatin structure and alter gene expression patterns in a human disease.**
- 85 Cai S, Han HJ, Kohwi-Shigematsu T: Tissue-specific nuclear architecture and gene expression regulated by SATB1. *Nat. Genet.* 34(1), 42–51 (2003).
- ■ **Demonstrates how deletion of polymorphic DNA repeats can change chromatin structure and alter gene expression patterns in a human disease.**
- 86 Yasui D, Miyano M, Cai S, Varga-Weisz P, Kohwi-Shigematsu T: SATB1 targets chromatin remodelling to regulate genes over long distances. *Nature* 419(6907), 641–645 (2002).
- 87 Notani D, Gottimukkala KP, Jayani RS *et al.*: Global regulator SATB1 recruits β -catenin and regulates T(H)2 differentiation in Wnt-dependent manner. *PLoS Biol.* 8(1), E1000296 (2010).
- 88 Pomerantz MM, Ahmadiyeh N, Jia L *et al.*: The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. *Nat. Genet.* 41(8), 882–884 (2009).
- 89 Pirozhkova I, Petrov A, Dmitriev P, Laoudj D, Lipinski M, Vassetzky Y: A functional role for 4qA/B in the structural rearrangement of the 4q35 region and in the regulation of *FRG1* and *ANT1* in facioscapulohumeral dystrophy. *PLoS One* 3(10), E3389 (2008).
- 90 Conrad DF, Pinto D, Redon R *et al.*: Origins and functional impact of copy number variation in the human genome. *Nature* 464(7289), 704–712 (2010).
- 91 Feuk L, Carson AR, Scherer SW: Structural variation in the human genome. *Nat. Rev. Genet.* 7(2), 85–97 (2006).
- 92 Feuk L, Marshall CR, Wintle RF, Scherer SW: Structural variants: changing the landscape of chromosomes and design of disease studies. *Hum. Mol. Genet.* 15(1), R57–R66 (2006).
- 93 Freeman JL, Perry GH, Feuk L *et al.*: Copy number variation: new insights in genome diversity. *Genome Res.* 16(8), 949–961 (2006).
- 94 Glessner JT, Wang K, Cai G *et al.*: Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 459(7246), 569–573 (2009).
- 95 Merikangas AK, Corvin AP, Gallagher L: Copy-number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet.* 25(12), 536–544 (2009).
- 96 Bassett AS, Marshall CR, Lionel AC, Chow EW, Scherer SW: Copy number variations and risk for schizophrenia in 22q11.2 deletion syndrome. *Hum. Mol. Genet.* 17(24), 4045–4053 (2008).
- 97 Han HJ, Russo J, Kohwi Y, Kohwi-Shigematsu T: SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* 452(7184), 187–193 (2008).
- 98 Simonis M, Klous P, Homminga I *et al.*: High-resolution identification of balanced and complex chromosomal rearrangements by 4C technology. *Nat. Meth.* 6(11), 837–842 (2009).
- 99 Dostie J, Zhan Y, Dekker J: Chromosome conformation capture carbon copy technology. *Curr. Protoc. Mol. Biol.* Chapter 21(Unit 21), 14 (2007).

- 100 Li G, Fullwood M, Xu H *et al.*: ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. *Genome Biol.* 11(2), R22 (2010).
- 101 Chambeyron S, Bickmore WA: Chromatin decondensation and nuclear reorganization of the HOXB locus upon induction of transcription. *Genes Dev.* 18(10), 1119–1130 (2004).
- 102 Sieben VJ, Marun CSD, Pilarski PM, Kaigala GV, Pilarski LM, Backhouse CJ: FISH and chips: chromosomal analysis on microfluidic platforms. *IET Nanobiotechnol.* 1(3), 27–35 (2007).
- 103 Yue P, Moulton J: Identification and analysis of deleterious human SNPs. *J. Mol. Biol.* 356(5), 1263–1274 (2006).
- 104 Kruglyak L, Nickerson DA: Variation is the spice of life. *Nat. Genet.* 27(3), 234–236 (2001).
- 105 International HapMap Consortium *et al.*: A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449(7164), 851–861 (2007).
- 106 Wain LV, Armour JAL, Tobin MD: Genomic copy number variation, human health, and disease. *Lancet* 374(9686), 340–350 (2009).
- 107 Mills RE, Luttig CT, Larkins CE *et al.*: An initial map of insertion and deletion (indel) variation in the human genome. *Genome Res.* 16(9), 1182–1190 (2006).
- 108 De La Chaux N, Messer P, Arndt P: DNA indels in coding regions reveal selective constraints on protein evolution in the human lineage. *BMC Evolutionary Biology* 7(1), 191 (2007).
- 109 Tuzun E, Sharp AJ, Bailey JA *et al.*: Fine-scale structural variation of the human genome. *Nat. Genet.* 37(7), 727–732 (2005).