

6-15-2010

Integrating one-dimensional and three-dimensional maps of genomes

Natalia Naumova

University of Massachusetts Medical School, Natalia.Naumova@umassmed.edu

Job Dekker

University of Massachusetts Medical School, Job.Dekker@umassmed.edu

Follow this and additional works at: http://escholarship.umassmed.edu/pgfe_pp



Part of the [Genetics and Genomics Commons](#)

Repository Citation

Naumova, Natalia and Dekker, Job, "Integrating one-dimensional and three-dimensional maps of genomes" (2010). *Program in Gene Function and Expression Publications and Presentations*. 57.
http://escholarship.umassmed.edu/pgfe_pp/57

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Program in Gene Function and Expression Publications and Presentations by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Integrating one-dimensional and three-dimensional maps of genomes

Natalia Naumova and Job Dekker*

Program in Gene Function and Expression and Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 364 Plantation Street, Worcester, MA, 01605-0103, USA

*Author for correspondence (job.dekker@umassmed.edu)

Journal of Cell Science 123, 1979-1988

© 2010. Published by The Company of Biologists Ltd

doi:10.1242/jcs.051631

Summary

Genomes exist in vivo as complex physical structures, and their functional output (i.e. the gene expression profile of a cell) is related to their spatial organization inside the nucleus as well as to local chromatin status. Chromatin modifications and chromosome conformation are distinct in different tissues and cell types, which corresponds closely with the diversity in gene-expression patterns found in different tissues of the body. The biological processes and mechanisms driving these general correlations are currently the topic of intense study. An emerging theme is that genome compartmentalization – both along the linear length of chromosomes, and in three dimensions by the spatial colocalization of chromatin domains and genomic loci from across the genome – is a crucial parameter in regulating genome expression. In this Commentary, we propose that a full understanding of genome regulation requires integrating three different types of data: first, one-dimensional data regarding the state of local chromatin – such as patterns of protein binding along chromosomes; second, three-dimensional data that describe the population-averaged folding of chromatin inside cells and; third, single-cell observations of three-dimensional spatial colocalization of genetic loci and trans factors that reveal information about their dynamics and frequency of colocalization.

This article is part of a Minifocus on exploring the nucleus. For further reading, please see related articles: ‘The nuclear envelope at a glance’ by Katherine L. Wilson and Jason M. Berk (*J. Cell Sci.* **123**, 1973-1978) and ‘Connecting the transcription site to the nuclear pore: a multi-tether process regulating gene expression’ by Guennaëlle Dieppois and Françoise Stutz (*J. Cell Sci.* **123**, 1989-1999).

Key words: Nuclear organization, Long-range gene regulation, Genome structure

Introduction

Many different approaches are being used to map the structural and functional state of chromosomes, and their relationship to gene regulation. The first is being used by a large number of studies that aim to map chromatin features along the linear genome (e.g. Barrera and Ren, 2006; Bernstein et al., 2007; Hon et al., 2009; Rando and Chang, 2009). Here, we refer to these as one-dimensional (1D) approaches. These studies involve mapping genes and their expression, the positions of histone modifications, DNase-I-hypersensitive sites, patterns of DNA methylation and binding of transcription factors (Box 1). The state of the art of these studies is perhaps best exemplified by the efforts of the ENCODE consortium, which is in the process of generating comprehensive 1D maps of the human genome in a panel of model cell lines (ENCODE-consortium, 2004; ENCODE-consortium, 2007). The second, more recently emerging set of approaches aims to analyze the spatial arrangement of chromosomes (Cremer et al., 2001; Fraser and Bickmore, 2007; Misteli, 2007; Zhao et al., 2009). These three-dimensional (3D) studies employ what are known as 3C-based methods to determine the population-averaged spatial proximity of distant genomic loci (Dekker, 2008; Dekker et al., 2002; Simonis et al., 2007), as well as high-resolution single-cell 3D imaging techniques to determine the sub-nuclear localization of loci with respect to each other and to (sub-) nuclear structures including nucleoli, nuclear bodies (such as transcription factories) and the nuclear periphery (see Box 2).

1D and 3D studies yield distinct but complementary descriptions of the same genome. Here, we propose that a unified and coherent understanding of genome regulation can only

be obtained by integrating 1D and 3D data. Integrating information obtained using these approaches is complicated because currently we do not fully understand the structural and functional relationships between 1D and 3D observations. 1D elements, such as genes and regulatory elements, can appear unrelated as judged by the fact that they are widely separated along the linear genome or even located on different chromosomes. However, 3D analysis can reveal that some of these regulatory elements are engaged in long-range looping interactions with target genes, or that groups of genes spread out throughout the genome congregate at the same sub-nuclear structures. Major emerging questions in this area of inquiry are: (1) What are the structural and functional relationships between the linear positions of genes and regulatory elements, and their spatial disposition? (2) What are the causal and mechanistic relationships between local (1D) chromatin state and 3D chromosome conformation and nuclear organization? (3) Do chromosome conformation, and nuclear organization affect gene expression, or vice versa (or both)? In this Commentary, we summarize emerging insights into how the genome is organized on the basis of 1D and 3D studies. In addition, we propose a path towards a more integrative analysis of chromosome biology that combines 1D and 3D approaches with single-cell observations that will begin to provide answers to some of these questions.

1D genome organization

Linear compartmentalization of chromosomes

Chromosomes are linear entities and, thus, are linearly organized. This is clearly illustrated by Giemsa staining of metaphase

Box 1. Methods of genome analysis in 1D

Transcription profiling is used to identify transcribed sections of a genome in a particular cell type. RNA is extracted, amplified and used to hybridize tiling microarrays or is directly sequenced.

Replication profiling is the measurement of replication timing along the genome. Replication timing informs on the positions of replication origins, replication-fork progression and convergence of replication forks. Replication timing can be determined using synchronized cells sorted at different stages of S phase. DNA extracted from these cells is then hybridized to microarrays or is analyzed by deep sequencing (Hansen et al., 2010; Hiratani et al., 2008; Jeon et al., 2005).

ChIP-chip or **ChIP-sequencing** are techniques that combine chromatin immunoprecipitation (ChIP) with microarray technology (chip) or deep sequencing (Johnson et al., 2007; Ren et al., 2000). These methods determine genomic locations at which specific proteins (e.g. RNA polymerase II, transcription factors) bind, or where specific protein modifications (e.g. histone modifications) occur. The method relies on formaldehyde crosslinking of proteins to DNA, followed by shearing and immunoprecipitation of specific protein-DNA complexes. DNA is then analyzed using microarrays or deep sequencing. Data analysis allows identification of promoters and regulatory elements, such as enhancers, silencers, insulators and locus control regions.

DamID is a powerful methylation-based in-vivo approach for mapping genomic interactions for a particular transcription factor or chromatin protein (van Steensel and Henikoff, 2000). The protein of interest is fused to *E. coli* DNA adenine methyltransferase (dam) and the fusion protein is expressed at very low levels in the cells. The fusion protein methylates adenines in DNA sequences surrounding native binding sites or distal sites that interact with the primary binding site. Adenine-methylated DNA fragments are isolated by selective PCR and can then be identified by microarray hybridization or deep sequencing. One advantage of this method over ChIP-based methods is that DamID does not require the use of specific antibodies or DNA crosslinking.

DNase I chip or **DNase I sequencing** are methods that detect DNase-I-hypersensitive sites (sites that are more readily cleaved by DNase I and that represent open chromatin) across genomes. As DNase I hypersensitivity is a universal feature of active cis-regulatory sequences, mapping of DNase I hypersensitivity is an accurate way of identifying promoter regions, enhancers and silencers, locus control regions and new elements. The method involves treatment of intact nuclei with DNase I, followed by detection of DNA fragment ends by ligation-mediated PCR amplification. Amplified DNA is then hybridized to tiling microarrays or is directly sequenced (Boyle et al., 2008; Crawford et al., 2006; Sabo et al., 2006).

chromosomes, which reveals the classic pattern of light and dark bands – so-called R-bands and G-bands (Bernardi, 2000; Furey and Haussler, 2003; Holmquist, 1992) (Fig. 1, top panel; Fig. 2A). These patterns are highly characteristic of specific chromosomes, independent of the cell type, and are a striking illustration of the distinct genomic features that are arranged in a linear (1D) manner along chromosomes. G-bands correspond to AT-rich and gene-poor segments of chromosomes, whereas R-bands are more GC-rich and gene-rich. Fly polytene chromosomes provide another visual example of structurally distinct domains along the length of chromosomes. DNA staining of these chromosomes produces a pattern of bands and interbands that correspond to condensed and more-open chromatin compartments, respectively (Zhimulev et al.,

2004). Thus, chromosomes are generally compartmentalized along their linear length, with alternating domains that differ in gene density and local chromatin state.

A variety of powerful high-resolution 1D approaches have been developed and are being applied to study chromosome organization (Fig. 1, Box 1). These methods map local chromatin features, as well as transcription and replication along the genome sequence, and provide detailed descriptions of the molecular features of chromosomal compartments. First, fractionation of the genome in open and more-compact fractions, followed by microarray analyses, shows that gene-dense and gene-poor compartments differ with respect to their levels of chromatin compaction (Gilbert et al., 2004). Second, analyses of transcription profiles indicate that highly expressed genes are clustered in specific chromosomal regions, referred to as regions of increased gene expression (RIDGES), that are related to classical R-bands (Caron et al., 2001; Kosak et al., 2007; Lercher et al., 2002; Lercher et al., 2003; Versteeg et al., 2003). Furthermore, analyses of patterns of histone modifications using chromatin immunoprecipitation (ChIP) combined with deep sequencing again reveal the presence of these same compartments. For example, ‘active’ histone marks, such as acetylation, are enriched in gene-dense compartments (Barski et al., 2007; Martens et al., 2005).

The domains described above tend to be large, typically several megabases in size, but 1D compartmentalization can also be discerned at a finer scale. Analysis of a variety of 1D datasets – including those that measured gene expression, DNase I hypersensitivity, timing of DNA replication and status of histone modification – that were generated for 1% of the human genome by the ENCODE pilot project showed the presence of chromatin domains that range from 20 Kb to 1 Mb (Thurman et al., 2006) (Fig. 1). Again, two types of domain were found: one contained most of the genes and displayed chromatin features that are typically associated with active chromatin, such as histone acetylation, open chromatin and early replication; the other was generally inactive or repressed.

Several observations suggest that compartmentalization of chromosomes is functionally relevant. First, gene expression analyses revealed that genes that are located near each other in the genome tend to be co-expressed, even when they are apparently functionally unrelated (Hurst et al., 2004). This phenomenon has been observed in *Saccharomyces cerevisiae* (Cohen et al., 2000), *Drosophila melanogaster* (de Wit et al., 2008; Spellman and Rubin, 2002), mouse (Kosak et al., 2007) and human (Caron et al., 2001; Lercher et al., 2002). The mechanisms that drive co-expression are not well understood, but it has been proposed that the generally open-chromatin state of gene-rich chromosomal compartments could poise resident genes for expression. This model is supported, for instance, by experiments in which a GFP-reporter gene is inserted at random locations in the mouse genome. Expression of the reporter was found to be related to the general expression status of the chromosomal domain into which it had been inserted (Gierman et al., 2007). Second, functionally related genes are often located next to one another in the linear genome. Well-studied examples include the gene clusters of α - and β -globin, and the Hox genes. Although clustering of these genes probably reflects the evolutionary process by which they were formed (e.g. gene duplication), clustering can be crucial for their coordinated regulation during development, as it can facilitate the sharing of common sets of nearby regulatory elements (de Laat and

Box 2. Methods of genome analysis in 3D**Chromosome conformation capture (3C)-based methods**

3C-based methods allow determination of the probabilities that genomic loci interact in nuclear space (Dekker et al., 2002). Chromosome conformation is captured by formaldehyde fixation, followed by digestion by restriction enzyme. Subsequently, ligation is performed under very dilute conditions that favor intramolecular ligation; thus, loci that are physically touching are converted into ligation products. The resulting library of 3C-ligation products represents the sum of long-range chromosomal interactions that occur throughout the genome. The various 3C-based methods differ mainly in the way the library of 3C-ligation products is interrogated (Simonis et al., 2007), but all of them capture interactions across all the cells of a population, thus providing population-averaged data.

3C involves analyzing the interactions between pairs of loci one by one through quantitative PCR, using specific primers to detect and quantify the corresponding ligation products in the library of 3C-ligation products (Dekker et al., 2002). 3C is typically used to analyze relatively small genomic regions (up to several hundred Kb) (Gheldof et al., 2006; Miele et al., 2009; Tolhuis et al., 2002).

4C (3C-on-chip or Circular 3C) allows detection of a genome-wide interaction profile of a single locus of interest (Simonis et al., 2006; Zhao et al., 2006). The basic notion of 4C is that most 3C ligation products are already circular, or can be circularized by digesting the 3C library using a frequent-cutting restriction enzyme, and then re-circularized by ligation. Therefore, 3C ligation products that involve a single locus of interest can be selectively amplified by inverse PCR. The PCR library is analyzed by microarray or deep sequencing.

5C allows the parallel detection of up to millions of interactions between two large sets of loci (Dostie and Dekker, 2007; Dostie et al., 2006). 5C employs ligation-mediated amplification with pools of 5C primers that are designed to anneal across 3C ligation junctions. The resulting 5C library is analyzed by microarray or deep sequencing. 5C can be used for interrogating interactions throughout a genomic region (or whole chromosome) to gain insights into its 3D folding, or for analysis of long-range interactions

between specific sets of functional elements (e.g. between a set of promoters and a set of enhancers) (Lajoie et al., 2009).

Hi-C is a truly unbiased method for genome-wide chromatin interaction discovery (Lieberman-Aiden et al., 2009). Hi-C data is used to map the 3D architecture of complete genomes. Hi-C is a 3C-based method in which ligation junctions are marked with biotin. DNA is then sheared and precipitated with streptavidin beads to selectively isolate DNA fragments with biotin-containing ligation junctions. The resulting Hi-C library is analyzed by deep sequencing.

Combinations of 3C-based methods and ChIP have also been developed (e.g. ChIP-loop, 6C, ChIA-PET, e4C). Here, chromatin is digested and then precipitated with an antibody in order to enrich for chromatin bound by the protein of interest. Subsequently, chromatin is intramolecularly ligated. The resulting ligation-product library is enriched in specific genetic elements with which the protein of interest associates through long-range chromatin interactions (Cai et al., 2006; Fullwood et al., 2009; Schoenfelder et al., 2010; Tiwari et al., 2008a).

Single-cell cytological techniques

Immunofluorescence with fluorescent antibodies is used to analyze the locations of specific proteins inside nuclei. This method is widely used for the detection of nuclear speckles (transcription factors, splicing machinery etc.) and transcription factories. In combination with RNA and DNA FISH, immunofluorescence is used to determine the colocalization of a locus of interest with nuclear proteins and substructures.

DNA fluorescence in situ hybridization (FISH) requires fixation, and DNA denaturation, and is used to visualize the spatial relationships between loci (or chromosomes) in single cells.

RNA FISH is used to reveal subnuclear localization of specific RNA molecules, and differs from DNA FISH in that it lacks a denaturation step. Sites of transcription can also be detected by growing cells in the presence of BrU. Incorporation of BrU shows – in a nonspecific manner – all sites of active transcription in living nuclei, and can be used to reveal transcription factories.

Grosveld, 2003; Sproul et al., 2005). For instance, the gene domain of β -globin contains a cluster of linked developmental- and differentiation-stage-specific genes (Stamatoyannopoulos, 2005). In erythroid cells, when the genes in the β -globin cluster are active, they form a domain of open chromatin (Bulger et al., 2003); the genes in this domain share common *cis*-regulatory sequences and replicate early during S phase in erythroid cells (Kitsberg et al., 1993). In non-erythroid cells, these genes are repressed, lie within compacted chromatin and replicate late during S phase (Kitsberg et al., 1993).

Therefore, it is clear that, at the level of chromosomal subdomains or groups of genes – i.e. at the scale of hundreds of Kb up to megabases – linear relationships along chromosomes are correlated with gene expression.

Genes and their regulatory elements can be far apart and out of linear order

Genes can be regulated by multiple regulatory elements. It is generally believed that regulatory elements are located on the same chromosome as their target gene (*cis*), although cases have been described in which elements on one chromosome affect genes on other chromosomes (*trans*) (Lomvardas et al., 2006; Spilianakis et al., 2005). In many cases, regulatory elements are located near their target genes (i.e. from a few Kb to 1 Mb away), as in the globin and the *Hox* gene clusters (Spitz et al., 2003; Stamatoyannopoulos, 2005). Thus, from a global point of view,

chromosomes appear to be composed of a linear series of genes, each directly surrounded by its regulatory elements.

Now that regulatory elements are being mapped with increasing accuracy and throughput (ENCODE-consortium, 2007), it is becoming clear that the number of regulatory elements far exceeds the number of genes, and that regulatory elements might be located at large genomic distances from their target genes. For instance, DNase-I-hypersensitive sites – the hallmark of a wide variety of regulatory elements such as promoters, enhancers and insulators – are abundant near genes but can also be found in large intergenic regions (Boyle et al., 2008; Crawford et al., 2006; Crawford et al., 2004; Sabo et al., 2006). Furthermore, regulatory elements are being identified in ChIP experiments that are used to identify sites bound by specific transcription factors. For instance, most of the mapped sites bound by the estrogen receptor (ER) are located quite far away (from tens up to hundreds of Kb) from promoters (Bourdeau et al., 2004; Carroll et al., 2006; Fullwood et al., 2009; Lin et al., 2007; Liu et al., 2008). Similar results were obtained in studies of the transcription factor GATA1 (Cheng et al., 2009; Fujiwara et al., 2009; Yu et al., 2009): this erythroid-specific transcriptional regulator binds sites that, in almost 50% of cases, are located further than 10 Kb from the nearest putative target gene. These studies imply that long-range phenomena, in which regulatory elements must act over large genomic distances (up to hundreds of Kb or more) to regulate genes, must be quite prevalent (Miele and Dekker, 2008).

There are over 900 transcription factors encoded by the *C. elegans* genome (Reece-Hoyes et al., 2005), and more than 1000 by the mouse and human genomes (Vaquerizas et al., 2009). Mapping of the binding sites of these factors is only slowly progressing, however, mainly owing to the lack of suitable antibodies. As more data are generated through ChIP studies and complementary gene-centered approaches such as the yeast one-hybrid system (Deplancke et al., 2006; Walhout, 2006), it is possible that we will learn that most transcription factors bind to sites all along chromosomes, often far from their target genes. The fact that regulatory elements can act over large genomic distances makes it conceivable that unrelated genes and elements could separate regulatory elements from their target genes. Indeed, detailed studies of the transcriptional regulation of individual loci have revealed that the linear order of genes and regulatory elements is not

necessarily directly related to functional relationships. That is, an element that regulates a given gene might be located within another unrelated gene or might regulate a gene located several hundreds of Kb away, but not a gene located nearby (e.g. Lettice et al., 2003; Spilianakis and Flavell, 2004) (for reviews, see Kleinjan and van Heyningen, 2005; West and Fraser, 2005). Our knowledge of such long-range relationships is still incomplete, and it is often difficult to assign target genes to known enhancers on the basis of only 1D annotations. In fact, in most cases, the relationships between regulatory elements and their target genes can only be indirectly inferred on the basis of correlations found between transcription factor binding and the analysis of gene expression [e.g. as in the case of ER-responsive genes (Carroll et al., 2006)].

Therefore, 1D analyses of chromosomes reveal that, at the megabase scale, chromosomes are linearly compartmentalized into

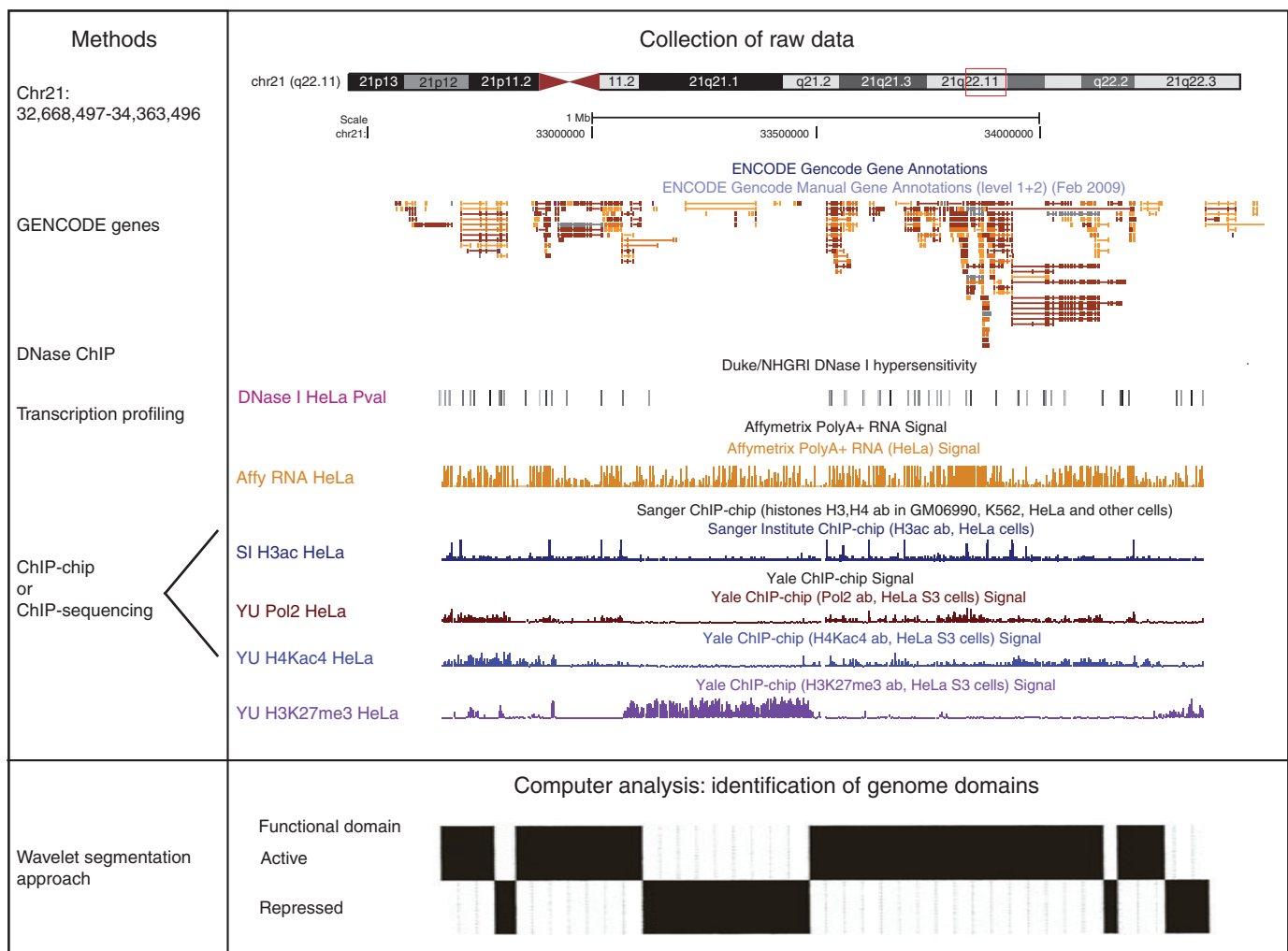


Fig. 1. 1D genome analysis. Mapping various genomic and chromatin features along chromosomes yields 1D data. The top part of the figure shows a screenshot of the Genome browser (<http://genome.ucsc.edu/>), presenting genome tracks for the ENCODE region Enm005. All data were generated by the ENCODE pilot project and are publicly available through the UCSC genome browser (ENCODE-consortium, 2007). Methods used to collect data are indicated on the left. Tracks indicate the presence of genes (GENCODE genes); the presence of DNase-I-hypersensitive sites (*P*-values are from HeLa cells; data from the ENCODE group at Duke University, led by Greg Crawford); the level of RNA (expression profiling data from HeLa cells; generated by the ENCODE group at Affymetrix, led by Tom Gingeras); the level of acetylated H3 (H3Ac; generated by ChIP-chip by the ENCODE group at the Sanger Institute, led by Ian Dunham); and levels of RNA polymerase II (Pol2) binding, and levels of acetylated H4 (H4Ac) and methylated H3 (H3K27me3) in HeLa cells as determined by ChIP-chip (data from the ENCODE group at Yale University, led by Michael Snyder). The bottom part of the figure shows an analysis of chromosomal domains, by using a combination of wavelet analysis and Hidden-Markov-Model segmentation of 1D datasets, including replication timing, DNase I hypersensitivity and histone modifications (adapted with permission from Thurman et al., 2006); see text for details.

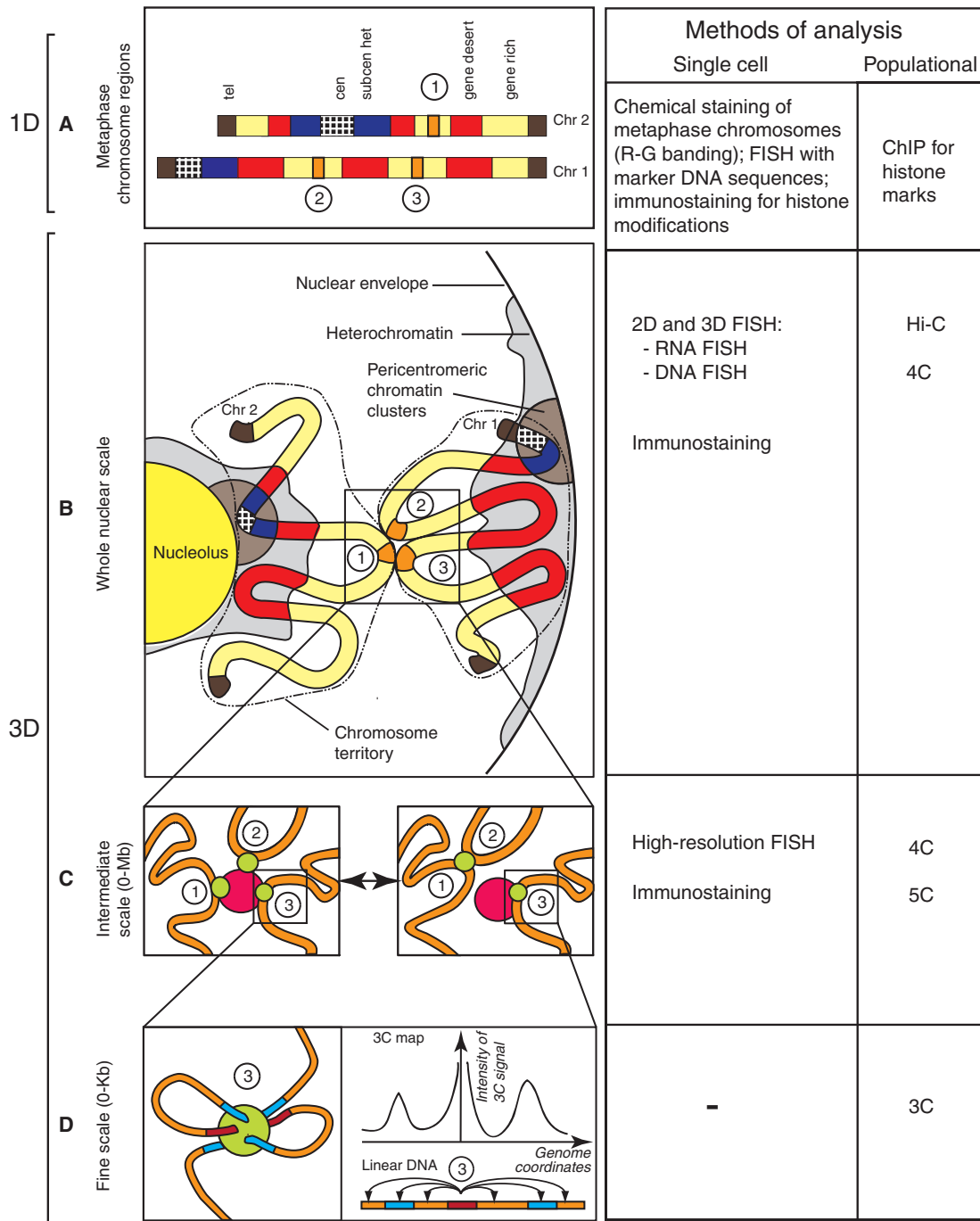


Fig. 2. 3D genome analysis. The spatial organization of genomes can be studied using single-cell methods or using population-based methods, and at different resolutions or length scales (A-D). (A) A hypothetical pair of metaphase chromosomes. 1D compartmentalization is indicated: constitutive heterochromatin domains include the centromere (cen), pericentromeric heterochromatin (subcen het) and telomeres (tel). Chromosome arms further consist of alternating active and repressed domains (indicated by different colors). Numbers indicate (chromosomal) regions to be analysed by 3D methods in panels B, C and D. (B) Spatial organization of chromosomes shown in A in the interphase nucleus. Chromosomal regions that are located far apart on the same chromosome (2 and 3) or located on different chromosomes (1 and 2, 1 and 3) can colocalize in 3D to form spatial compartments. (C) A higher-resolution (Mb scale) analysis of *cis* and *trans* associations of chromosomal regions 1 to 3 shown in B. At this resolution, associations of groups of genes can be detected surrounding subnuclear structures such as transcription factories (green circles) and splicing bodies [characterized by the presence of the splicing factor SC35 (red)]. For example, a *trans* interaction between regions 1 and 2 can occur through colocalization to the same transcription factory (Osborne et al., 2004) or to two different transcription factories that are both associated with one SC35 granule (Brown et al., 2008). (D) High-resolution (Kb scale) analysis of 3D folding and long-range associations that can be studied using 3C-based methods. At this scale, specific looping interactions can be detected between genes and regulatory elements. This scheme provides an example of a 3C analysis in which the interaction probability of a single defined genomic element is mapped throughout the larger region 3 (right). Peaks in this 3C map indicate long-range interactions that suggest a looped conformation (indicated on the left). Visual representation of the chromosomes and nucleus were inspired by Fig. 4J of the article published by Solovei and colleagues (Solovei et al., 2009).

functionally distinct sub-chromosomal domains. At a finer scale – i.e. at the level of genes and regulatory elements – the correlation between position and function is less clear, because the linear relationship between genetic loci and their regulatory elements cannot necessarily be used to predict gene expression.

3D genome organization

Spatial compartmentalization of the genome

Chromosomes are not simply 1D entities – their 3D organization inside the nucleus can have important roles in genome regulation (Dekker, 2008; Fraser and Bickmore, 2007; Miele and Dekker, 2008; Misteli, 2007). The 3D organization of chromosomes – and the nucleus in general – has traditionally been studied by microscopic methods. More recently, molecular methods based on the chromosome conformation capture technique (3C) (Dekker et al., 2002) have become widely applied for analysis of long-range chromatin interactions between genomic loci. Now that these methods allow genome-wide mapping of long-range interactions (e.g. the Hi-C method, which analyses spatially adjacent DNA across the entire genome), complete 3D views of the genome can be obtained (Lieberman-Aiden et al., 2009). These new high-throughput and high-resolution methods have already led to new insights, including the notion that long-range associations between genomic loci are a commonly employed control mechanism of gene expression, and that they have crucial roles in determining nuclear organization in general. Box 2 summarizes these methods.

At the level of the whole nucleus, chromosomes form individual compartments, so-called chromosome territories (Fig. 2) (Cremer and Cremer, 2001). These can be directly observed by fluorescence in-situ hybridization (FISH) studies using whole-chromosome paint probes. The formation of territories implies that chromosomes do not readily mix with one another, although some level of intermingling does occur, which provides opportunities for *trans* interactions between loci (Branco and Pombo, 2006; Fraser, 2006). Furthermore, chromosomes have preferred, but not absolute, subnuclear positions; particular chromosomes are located more frequently at the nuclear periphery than at the center of the nucleus, whereas others display the opposite tendency and are more often located near the center of the nucleus (Cremer et al., 2001; Croft et al., 1999; Lieberman-Aiden et al., 2009).

Chromosome territories are also internally compartmentalized. Recent genome-wide chromatin-interaction maps of the human genome generated using Hi-C technology (Box 2) showed that chromosomes are distributed over two types of ‘spatial compartments’: one compartment is enriched in active genes and open chromatin (compartment A), whereas the other compartment contains inactive and closed chromatin (compartment B) (Lieberman-Aiden et al., 2009). Interestingly, genomic regions found in these spatial compartments closely correlate with gene-rich and gene-poor regions identified along the linear genome (described above). Data obtained through Hi-C, 4C and FISH approaches showed that active domains along a given chromosome tend to preferentially interact with each other – i.e. they are near each other in 3D despite being separated by large linear genomic distances (Brown et al., 2008; Iborra et al., 1996a; Lieberman-Aiden et al., 2009; Osborne et al., 2004; Schoenfelder et al., 2010; Simonis et al., 2006; Solovei et al., 2009) (Fig. 2B). However, inactive compartments also preferentially associate with one another (Bantignies et al., 2003; Lieberman-Aiden et al., 2009; Simonis et al., 2006). Thus, the alternating linear compartments defined by active and inactive local chromatin states also define two types of

spatial compartment in which chromosomal domains with similar chromatin state (i.e. active/open versus inactive/closed chromatin) come together (Fig. 2B). Interestingly, these spatial compartments are not restricted to single chromosomes and also contain regions of other chromosomes, as demonstrated by Hi-C, 4C and FISH analyses (Lieberman-Aiden et al., 2009; Osborne et al., 2004; Schoenfelder et al., 2010; Simonis et al., 2006; Zhao et al., 2006). Spatial clustering of groups of loci could be functionally relevant, similar to the functional consequences that result from the clustering of groups of genes along the linear genome sequence (see above). In one model, spatial clustering results in high local concentration of transcription (or silencing) machinery components, which could increase the efficiency of transcription (or silencing) (Miele et al., 2009; Sutherland and Bickmore, 2009).

As stated above, genes and their regulatory elements can be located far apart, and can even be in separate chromosomal domains (i.e. be located in separate 1D compartments). 3C-based studies have convincingly shown that regulatory elements can act over large genomic distances by chromatin looping (Cai et al., 2006; Fullwood et al., 2009; Kurukuti et al., 2006; Majumder et al., 2008; Spilianakis and Flavell, 2004; Tiwari et al., 2008b; Tolhuis et al., 2002; Vakoc et al., 2005; Vernimmen et al., 2007; Zhou et al., 2006). The role of long-range looping interactions in gene regulation has been extensively covered in a number of recent reviews (Chambeyron and Bickmore, 2004; de Laat and Grosfeld, 2003; Dekker, 2008; Fraser, 2006; Göndör and Ohlsson, 2009; Kadauke and Blobel, 2009; Miele and Dekker, 2008; Vilar and Saiz, 2005), and will not be treated here in detail. We speculate that the formation of the large spatial compartments that are enriched in active chromatin segments described above brings large sets of genes and regulatory elements located in different domains along the linear genome in general spatial proximity (mainly in *cis*, but also in *trans*). This would also facilitate the sometimes very long-range interactions between specific sets of regulatory elements and their target genes. Thus, as for large chromosomal domains, genes and regulatory elements that are widely spaced along the linear genome come together in 3D. In fact, the frequency of 3D colocalization might be a more accurate predictor of functional relationships between genes and regulatory elements than their 1D proximity along the linear genome.

Collectively, these observations suggest that spatial separation of active and inactive chromatin domains, as well as the clustering of functionally related loci, are general principles at all length scales: at the level of the nucleus, of whole chromosomes, of sub-chromosomal domains, and at the gene and gene-regulatory-element level. A major unanswered question revolves around the specificity (and functional relevance) of these associations at each of these levels. Although it seems probable that the interactions between genes and regulatory elements are in most cases specific to ensure appropriate spatiotemporal regulation of gene expression, the specificity of these interactions at the scale of subchromosomal domains and the whole nucleus is currently an unresolved issue.

Subnuclear localization of transcription and repression machineries, and their correlation with spatial chromosome compartments

Microscopy studies have revealed that processes such as transcription, splicing, DNA replication and repair, and repression of gene expression occur at subnuclear foci that are enriched in the relevant *trans*-acting protein machineries (Dellaire and Bazett-Jones, 2007; Iborra et al., 1996a; Kitamura et al., 2006). For

instance, sites of transcription can be visualized by 5-bromo-uridine (BrU) incorporation into nascent RNAs or by staining for active RNA polymerases (Sutherland and Bickmore, 2009). In addition, transcription mediated by all three nuclear RNA polymerases appears to be carried out at discrete sites that are specific for each type of polymerase (Iborra et al., 1996a; Pombo et al., 1999). These sites of transcription are sometimes referred to as 'transcription factories', although their composition and mode of formation are still controversial (Sutherland and Bickmore, 2009). More specifically, one model proposes that sites of transcription are pre-formed, and that genes migrate to these locations to be transcribed (Mitchell and Fraser, 2008; Iborra et al., 1996b); an alternate model states that active genes self-assemble to form transcription foci (Cook, 2002), perhaps through entropic forces (Cook and Marenduzzo, 2009) or by (weak) attractive forces between locally recruited protein complexes.

A well-known example of a specialized nuclear site of transcription is the nucleolus, a self-assembling structure of 200-500 nm in diameter where ribosomal genes are transcribed by RNA polymerase I. Most other genes are transcribed by RNA polymerase II. Punctate sites of gene transcription are found throughout the nucleus, and it is thought that, at each of these sites (transcription factories), multiple genes are being actively transcribed (Iborra et al., 1996a; Martin and Pombo, 2003; Osborne et al., 2004). Similarly, proteins involved in gene silencing are also enriched in other subnuclear locations. One example is the formation of polycomb bodies at specific sites where polycomb-repressed loci cluster together (Bantignies et al., 2003; Lanzaolo et al., 2007).

Therefore, the nucleus appears to be composed of a large number of small neighborhoods that are enriched in either active or inactive chromatin. These neighborhoods, which can be detected by immunofluorescence, might correspond to the spatial compartments that have been detected using 3C-based assays (Lieberman-Aiden et al., 2009; Schoenfelder et al., 2010; Simonis et al., 2006). For example, a recent study by the Fraser laboratory demonstrated that groups of expressed genes that were found to physically associate (as determined by e4C, a 3C-based assay) were also observed (by FISH) to be colocalized at transcription factories (Schoenfelder et al., 2010).

Are all active (repressed) neighborhoods equal?

The notion that the nucleus is composed of a large number of active or inactive neighborhoods immediately raises the question of whether there is any specificity in the composition of these neighborhoods. In particular, do functionally related genes – i.e. those regulated by a similar set of transcription factors – preferentially associate to form specialized active neighborhoods (or transcription factories)? Observations that both support and oppose such a model have been reported. For instance, in support of this model, direct evidence of specialization of transcription factories comes from studies of subnuclear localization when different plasmids were transfected into COS7 cells (Xu and Cook, 2008): plasmids carrying the same promoter were more likely to cluster together than plasmids containing different promoters. Notably, selective association between the plasmids was also found to depend on the presence of an intron in the genes on both plasmids. This finding suggests that association with transcription factories is also linked to mRNA splicing.

There is also evidence that functionally related endogenous genes associate. For instance, Schoenfelder and co-workers found that the active β -globin locus associated with a large number of active genes in erythroid cells (Schoenfelder et al., 2010). In particular,

genes regulated by the erythroid-specific transcription factor Krüppel-like factor 1 (KLF1) were found to associate with the β -globin locus, and these associations occurred at transcription factories enriched in KLF1. These findings suggest the presence of specialized transcription sites inside the nucleus where co-regulated and functionally related genes are clustered together. Another example is the close spatial association of the interferon γ (*Irfng*) gene on chromosome 10 with regulatory elements that control the expression of the interleukin (IL) gene cluster containing *Il4*, *Il5* and *Il13* on chromosome 11. It has been proposed that this association between *Irfng* and the regulatory elements of the IL gene cluster is important for coordinating mutually exclusive expression of these genes (Spilianakis et al., 2005). Further indications that co-regulated loci cluster together come from studies showing that many transcription factors are localized to subnuclear foci. For instance, steroid hormone receptors form 'speckles' upon binding their activating ligands; it is probable that the genes targeted by these nuclear receptors are also localized at these sites. Consistently, Fullwood and co-workers reported that multiple ER-bound sites often interact with each other to form looped clusters that are involved in regulation of ER target genes (Fullwood et al., 2009).

Conversely, other studies suggest a lack of association of specific genes at specialized transcription factories. For example, the de Laat laboratory used 4C to analyze associations between the mouse β -globin locus and the rest of the genome, and showed that the active locus associates with other active genes throughout the genome, whereas the inactive locus is more frequently associated with other inactive genes (Simonis et al., 2006). However, the authors did not find that the active locus preferentially associates with other functionally related erythroid-specific genes, suggesting that there is no, or only very limited, specificity or preference in these associations.

It is important to clarify that any preference of co-regulated and functionally related genes to associate in 3D space is probably not absolute. Both of the studies on the β -globin locus described above found that most of the genes that associate with active β -globin genes are not erythroid-specific genes or genes that are thought to share specific regulatory factors. Thus, our view is that the issue of whether there are specific associations among co-regulated and functionally related genes is not settled. Notably, in several cases, foci formed by transcription factors do not overlap with sites of target-gene transcription. For instance, foci containing the transcription factors OCT1 and E2F1, and the (steroid) glucocorticoid receptor, were not found to overlap with sites of BrU incorporation (Grande et al., 1997).

Population-averaged data versus single-cell observations

When data obtained by immunofluorescence (e.g. detection of subnuclear transcription sites) are compared with genome-wide interaction data (e.g. obtained by 4C and Hi-C approaches), an apparent discrepancy emerges. Genome-wide profiles of chromatin interactions show that a given gene interacts with a large number of loci throughout the genome. For instance, the active mouse β -globin locus interacts with a large number of other active genes (Schoenfelder et al., 2010; Simonis et al., 2006). However, immunofluorescence analysis of transcription sites suggests that only very few (eight to 20) genes are transcribed at a given subnuclear site or transcription factory (Jackson et al., 1998; Martin and Pombo, 2003), thereby indicating that, in any given nucleus, a gene can associate with or be located close to only a few other loci

or regulatory elements. The reason why these different approaches lead to different conclusions is because 3C-based chromatin-interaction studies typically analyze many millions of cells and, therefore, report on quantitative trends in a large population of cells. By contrast, imaging studies analyze individual cells in much smaller numbers, typically on the order of hundreds at most.

One model that unifies the data obtained by using both types of approach is the one that suggests that, at any given moment in a given cell, an active gene (or chromosomal locus in general) is associated with (or is located close to) only a limited number of other active genes (or loci), but that the precise composition of this group is different in every cell (e.g. Lomvardas et al., 2006; Sandhu et al., 2009). In other words, the genome-wide active and inactive spatial compartments identified by chromatin-interaction studies should actually be viewed as a large collection of much smaller neighborhoods that each contain only a few chromosomal segments. Furthermore, the precise composition of each neighborhood of active (or inactive) chromatin varies across a population of cells. Importantly, 3C-based assays clearly indicate that there are general tendencies for certain loci to be more frequently located close to each other than to others. Therefore, the spatial organization of chromosomes is probabilistic in nature, as has been previously proposed (de Laat and Grosfeld, 2007; Misteli, 2001). Population-based chromatin-interaction studies reveal the probability distributions of contacts within an entire cell population, whereas single-cell analyses provide an example of an individual set of associations in a given cell, thereby highlighting the variability in long-range associations between specific loci that occur from cell to cell.

Integrating 1D and 3D approaches

When 1D, 3D or single-cell studies are performed in isolation, each reveals only a limited view of the genome. Only when these data are combined can we interpret each set of data correctly, and obtain a full appreciation of the complexity of chromosome form and function. For instance, 1D approaches such as ChIP can identify the binding sites for a transcription factor along the genome (Fig. 3) but the dataset will not necessarily reveal the target genes of these putative regulatory elements. 3D approaches such as 3C or its derivations can be used to determine whether these elements loop to distant target genes, and to determine the colocalization of sets of genes. Finally, single-cell analysis by using microscopic methods can reveal a final layer of genome organization, by demonstrating that these loci associate in only in small groups and have compositions that vary in different cells (Fig. 3). Furthermore, only single-cell analysis will provide insights into the dynamics of individual spatial associations, for example, upon induction of transcription.

Besides complementing one another, these different approaches can also help to correctly interpret individual datasets. First, not all sites identified in a ChIP experiment are the primary binding site of the protein that is being studied; ChIP can also identify sites that are only indirectly bound by the protein through a chromatin-looping interaction (Fig. 3, asterisk). 3C-based assays can be used to determine whether such looping interactions occur and, if so, with which other protein-bound sites. Second, although large-scale chromatin interaction studies [e.g. by (e)4C, 5C, Hi-C and ChiA-PET approaches] can show the tendency of sets of loci to be located close to each other, only microscopic studies reveal the true frequency at which specific associations occur and the average spatial distance between loci. Third, 3D studies can reveal close

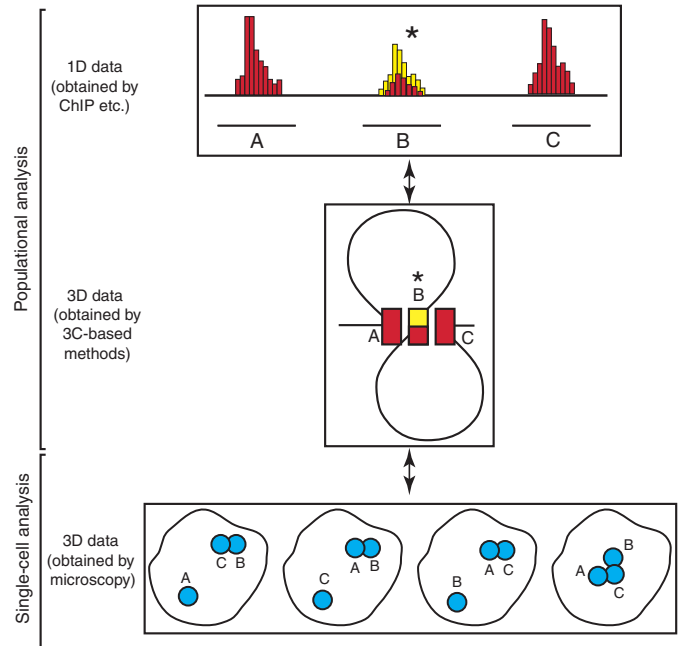


Fig. 3. Integration of 1D and 3D genome annotations. Data collected by 1D and 3D approaches supplement and complement one another. 1D genome annotations – for example, transcription factor (TF)-binding sites, obtained by ChIP-sequencing – indicate the locations of potential regulatory elements. (Top panel) Binding profiles for two TFs (red and yellow), which show association of TF red with regulatory elements A, B and C (red bars), and association of TF yellow with regulatory element B only (yellow bars). Note that these binding profiles do not identify the target genes controlled by these elements because the target genes can be located elsewhere in the genome. (Middle panel) 3D annotations, such as looping interactions between genes and elements detected by 3C-based techniques, can help to assign regulatory elements to their target genes. Here we show an example of long-range interactions between the three regulatory elements A, B and C indicated in the top panel. Red and yellow colors indicate which transcription factor was detected on each element by ChIP-sequencing. (Bottom panel) Both 1D data and 3C-based 3D data are obtained from large cell populations (millions) and represent the sum of all interactions that occur in that population. Single-cell 3D analysis (e.g. determining colocalization of loci by microscopic methodologies) shows the real quantitative spatial relationships between genomic regions. The bottom panel shows four individual nuclei that each have differential subnuclear localization of elements A, B and C, illustrating that different combinations of long-range interactions between loci occur in different cells. By combining the information from the three panels shown in this figure – i.e. by integrating 1D and 3D studies – a more comprehensive model of the functional state of the chromatin can be obtained. In this example, one can hypothesize that TF red binds to element B (asterisk, upper panel) only indirectly, through a looping interaction with elements A and/or C (see middle panel). Further single-cell analysis of colocalization of these elements and bound transcription factors (obtained by combining DNA FISH and immunofluorescence-mediated detection of proteins) can resolve this issue. For instance, if binding of TF red to element B – as detected by ChIP – is an indirect effect, single-cell observations should reveal that TF red is only found at element B when element B colocalizes with element A or C.

spatial relationships between sets of loci in specific locations of the nucleus (e.g. near the nuclear lamina, nuclear speckles etc.), which can suggest that they are co-regulated and share common regulatory proteins. 1D experiments, such as gene-expression profiling and ChIP for transcription factors, can be used to identify factors that are bound, and might determine the colocalization of these loci.

Future perspectives

The spatial organization of chromosomes promises to be the topic of intense study for years to come. Over the last several years, powerful new technologies have pushed the field forward. Molecular genome-wide methods, such as RNA-sequencing, ChIP-sequencing, 4C, 5C and Hi-C now enable detailed studies of local chromatin states and 3D folding of complete genomes. Sophisticated imaging methods allow a peek into how the genome is organized in single cells, in some cases in living cells in real time. Together, these methods can be used to address many open questions in the field. First, what are the mechanisms by which loci that are located far apart in the genome become spatially colocalized? To what extent is this spatial organization of the genome deterministic versus the result of self-organization? Which long-range associations and interactions are specific and functional? What is the benefit of formation of neighborhoods that are enriched in active and inactive loci? When 1D and 3D methods are integrated and combined with real-time single-cell studies, we will be in a better position to address these long-standing mechanistic questions.

The Dekker lab is supported by a grant from the National Institutes of Health (HG003143) and a W. M. Keck Foundation Distinguished Young Scholar Award (to J.D.). We thank Boris Joffe and Irina Solovoi for suggestions for Fig. 2. Deposited in PMC for release after 12 months.

References

- Bantignies, F., Grimaud, C., Lavrov, S., Gabut, M. and Cavalli, G. (2003). Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. *Genes Dev.* **17**, 2406-2420.
- Barrera, L. O. and Ren, B. (2006). The transcriptional regulatory code of eukaryotic cells—insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr. Opin. Cell Biol.* **18**, 291-298.
- Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I. and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823-837.
- Bernardi, G. (2000). Isochores and the evolutionary genomics of vertebrates. *Gene* **241**, 3-17.
- Bernstein, B. E., Meissner, A. and Lander, E. S. (2007). The mammalian epigenome. *Cell* **128**, 669-681.
- Bourdeau, V., Deschênes, J., Métivier, R., Nagai, Y., Nguyen, D., Bretschneider, N., Gannon, F., White, J. H. and Mader, S. (2004). Genome-wide identification of high-affinity estrogen response elements in human and mouse. *Mol. Endocrinol.* **18**, 1411-1427.
- Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S. and Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-322.
- Branco, M. R. and Pombo, A. (2006). Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, e138.
- Brown, J. M., Green, J., das Neves, R. P., Wallace, H. A., Smith, A. J., Hughes, J., Gray, N., Taylor, S., Wood, W. G., Higgs, D. R. et al. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083-1097.
- Bulger, M., Schubeler, D., Bender, M. A., Hamilton, J., Farrell, C. M., Hardison, R. C. and Groudine, M. (2003). A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. *Mol. Cell Biol.* **23**, 5234-5244.
- Cai, S., Lee, C. C. and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat. Genet.* **38**, 1278-1288.
- Caron, H., van Schaik, B., van der Mee, M., Baas, F., Riggins, G., van Sluis, P., Hermus, M. C., van Asperen, R., Boon, K., Voute, P. A. et al. (2001). The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289-1292.
- Carroll, J. S., Meyer, C. A., Song, J., Li, W., Geistlinger, T. R., Eeckhoutte, J., Brodsky, A. S., Keeton, E. K., Fertuck, K. C., Hall, G. F. et al. (2006). Genome-wide analysis of estrogen receptor binding sites. *Nat. Genet.* **38**, 1289-1297.
- Chambeyron, S. and Bickmore, W. A. (2004). Does looping and clustering in the nucleus regulate gene expression? *Curr. Opin. Cell Biol.* **16**, 256-262.
- Cheng, Y., Wu, W., Kumar, S. A., Yu, D., Deng, W., Triplic, T., King, D. C., Chen, K. B., Zhang, Y., Drautz, D. et al. (2009). Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res.* **19**, 2172-2184.
- Cohen, B. A., Mitra, R. D., Hughes, J. D. and Church, G. M. (2000). A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat. Genet.* **26**, 183-186.
- Cook, P. R. (2002). Predicting three-dimensional genome structure from transcriptional activity. *Nat. Genet.* **32**, 347-352.
- Cook, P. R. and Marenduzzo, D. (2009). Entropic organization of interphase chromosomes. *J. Cell Biol.* **186**, 825-834.
- Crawford, G. E., Holt, I. E., Mullikin, J. C., Tai, D., Blakesley, R., Bouffard, G., Young, A., Masiello, C., Green, E. D., Wolfsberg, T. G. et al. (2004). Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl. Acad. Sci. USA* **101**, 992-997.
- Crawford, G. E., Davis, S., Scacheri, P. C., Renaud, G., Halawi, M. J., Erdos, M. R., Green, R., Meltzer, P. S., Wolfsberg, T. G. and Collins, F. S. (2006). DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat. Methods* **3**, 503-509.
- Cremer, M., von Hase, J., Volm, T., Brero, A., Kreth, G., Walter, J., Fischer, C., Solovoi, I., Cremer, C. and Cremer, T. (2001). Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells. *Chromosome Res.* **9**, 541-567.
- Cremer, T. and Cremer, C. (2001). Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat. Rev. Genet.* **2**, 292-301.
- Croft, J. A., Bridger, J. M., Boyle, S., Perry, P., Teague, P. and Bickmore, W. A. (1999). Differences in the localization and morphology of chromosomes in the human nucleus. *J. Cell Biol.* **145**, 1119-1131.
- de Laat, W. and Grosveld, F. (2003). Spatial organization of gene expression: the active chromatin hub. *Chromosome Res.* **11**, 447-459.
- de Laat, W. and Grosveld, F. (2007). Inter-chromosomal gene regulation in the mammalian cell nucleus. *Curr. Opin. Genet. Dev.* **17**, 456-464.
- de Wit, E., Braunschweig, U., Greil, F., Bussemaker, H. J. and van Steensel, B. (2008). Global chromatin domain organization of the *Drosophila* genome. *PLoS Genet.* **4**, e1000045.
- Dekker, J. (2008). Gene regulation in the third dimension. *Science* **319**, 1793-1794.
- Dekker, J., Rippe, K., Dekker, M. and Kleckner, N. (2002). Capturing chromosome conformation. *Science* **295**, 1306-1311.
- Dellaire, G. and Bazett-Jones, D. P. (2007). Beyond repair foci: subnuclear domains and the cellular response to DNA damage. *Cell Cycle* **6**, 1864-1872.
- Deplancke, B., Mukhopadhyay, A., Ao, W., Elewa, A. M., Grove, C. A., Martinez, N. J., Sequerra, R., Doucette-Stamm, L., Reece-Hoyes, J. S., Hope, I. A. et al. (2006). A gene-centered C. elegans protein-DNA interaction network. *Cell* **125**, 1193-1205.
- Dostie, J. and Dekker, J. (2007). Mapping networks of physical interactions between genomic elements using 5C technology. *Nat. Protoc.* **2**, 988-1002.
- Dostie, J., Richmond, T. A., Arnaout, R. A., Selzer, R. R., Lee, W. L., Honan, T. A., Rubio, E. H., Krumm, A., Lamb, J., Nusbaum, C. et al. (2006). Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299-1309.
- ENCODE-consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-640.
- ENCODE-consortium (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816.
- Fraser, P. (2006). Transcriptional control thrown for a loop. *Curr. Opin. Genet. Dev.* **16**, 490-495.
- Fraser, P. and Bickmore, W. (2007). Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**, 413-417.
- Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A. K., Kang, Y. A., Choi, K., Farnham, P. J. and Bresnick, E. H. (2009). Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol. Cell* **36**, 667-681.
- Fullwood, M. J., Liu, M. H., Pan, Y. F., Liu, J., Xu, H., Mohamed, Y. B., Orlov, Y. L., Velkov, S., Ho, A., Mei, P. H. et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64.
- Furey, T. S. and Haussler, D. (2003). Integration of the cytogenetic map with the draft human genome sequence. *Hum. Mol. Genet.* **12**, 1037-1044.
- Gheldof, N., Tabuchi, T. M. and Dekker, J. (2006). The active FMR1 promoter is associated with a large domain of altered chromatin conformation with embedded local histone modifications. *Proc. Natl. Acad. Sci. USA* **103**, 12463-12468.
- Gierman, H. J., Indemans, M. H., Koster, J., Goetze, S., Seppen, J., Geerts, D., van Driel, R. and Versteeg, R. (2007). Domain-wide regulation of gene expression in the human genome. *Genome Res.* **17**, 1286-1295.
- Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N. P. and Bickmore, W. A. (2004). Chromatin architecture of the human genome; gene-rich domains are enriched in open chromatin fibers. *Cell* **118**, 555-566.
- Göndör, A. and Ohlsson, R. (2009). Chromosome crosstalk in three dimensions. *Nature* **461**, 212-217.
- Grande, M. A., van der Kraan, I., de Jong, L. and van Driel, R. (1997). Nuclear distribution of transcription factors in relation to sites of transcription and RNA polymerase II. *J. Cell Sci.* **110**, 1781-1791.
- Hansen, R. S., Thomas, S., Sandstrom, R., Canfield, T. K., Thurman, R. E., Weaver, M., Dorschner, M. O., Gartler, S. M. and Stamatoyannopoulos, J. A. (2010). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc. Natl. Acad. Sci. USA* **107**, 139-144.
- Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C. W., Lyou, Y., Townes, T. M., Schübeler, D. and Gilbert, D. M. (2008). Global reorganization of replication domains during embryonic stem cell differentiation. *PLoS Biol.* **6**, e245.
- Holmquist, G. P. (1992). Chromosome bands, their chromatin flavors, and their functional features. *Am. J. Hum. Genet.* **51**, 17-37.
- Hon, G. C., Hawkins, R. D. and Ren, B. (2009). Predictive chromatin signatures in the mammalian genome. *Hum. Mol. Genet.* **18**, R195-R201.
- Hurst, L. D., Pal, C. and Lercher, M. J. (2004). The evolutionary dynamics of eukaryotic gene order. *Nat. Rev. Genet.* **5**, 299-310.
- Iborra, F. J., Pombo, A., Jackson, D. A. and Cook, P. R. (1996a). Active RNA polymerases are localized within discrete transcription 'factories' in human nuclei. *J. Cell Sci.* **109**, 1427-1436.

- Iborra, F. J., Pombo, A., McManus, J., Jackson, D. A. and Cook, P. R. (1996b). The topology of transcription by immobilized polymerases. *Exp. Cell Res.* **229**, 167-173.
- Jackson, D. A., Iborra, F. J., Manders, E. M. and Cook, P. R. (1998). Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Mol. Biol. Cell* **9**, 1523-1536.
- Jeon, Y., Bekiranov, S., Karnani, N., Kapranov, P., Ghosh, S., MacAlpine, D., Lee, C., Hwang, D. S., Gingeras, T. R. and Dutta, A. (2005). Temporal profile of replication of human chromosomes. *Proc. Natl. Acad. Sci. USA* **102**, 6419-6424.
- Johnson, D. S., Mortazavi, A., Myers, R. M. and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-1502.
- Kadauke, S. and Blobel, G. A. (2009). Chromatin loops in gene regulation. *Biochim. Biophys. Acta.* **1789**, 17-25.
- Kitamura, E., Blow, J. J. and Tanaka, T. U. (2006). Live-cell imaging reveals replication of individual replicons in eukaryotic replication factories. *Cell* **125**, 1297-1308.
- Kitsberg, D., Selig, S., Keshet, I. and Cedar, H. (1993). Replication structure of the human beta-globin gene domain. *Nature* **366**, 588-590.
- Kleinjan, D. A. and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**, 8-32.
- Kosak, S. T., Scalzo, D., Alworth, S. V., Li, F., Palmer, S., Enver, T., Lee, J. S. and Groudine, M. (2007). Coordinate gene regulation during hematopoiesis is related to genomic organization. *PLoS Biol.* **5**, e309.
- Kurukuti, S., Tiwari, V. K., Tavossidana, G., Pugacheva, E., Murrell, A., Zhao, Z., Lobanenko, V., Reik, W. and Ohlsson, R. (2006). CTCF binding at the H19 imprinting control region mediates maternally inherited higher-order chromatin conformation to restrict enhancer access to Igf2. *Proc. Natl. Acad. Sci. USA* **103**, 10684-10689.
- Lajoie, B. R., van Berkum, N. L., Sanyal, A. and Dekker, J. (2009). MyS3C: web tools for chromosome conformation capture studies. *Nat. Methods* **6**, 690-691.
- Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F. and Orlando, V. (2007). Polycomb elements mediate the formation of chromosome higher order structures in the bithorax complex. *Nat. Cell Biol.* **9**, 1167-1174.
- Lercher, M. J., Urrutia, A. O. and Hurst, L. D. (2002). Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nat. Genet.* **31**, 180-183.
- Lercher, M. J., Urrutia, A. O., Pavlicek, A. and Hurst, L. D. (2003). A unification of mosaic structures in the human genome. *Hum. Mol. Genet.* **12**, 2411-2415.
- Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725-1735.
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, L., Ragozy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O. et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293.
- Lin, C. Y., Vega, V. B., Thomsen, J. S., Zhang, T., Kong, S. L., Xie, M., Chiu, K. P., Lipovich, L., Barnett, D. H., Stossi, F. et al. (2007). Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genet.* **3**, e87.
- Liu, Y., Gao, H., Marstrand, T. T., Ström, A., Valen, E., Sandelin, A., Gustafsson, J. A. and Dahlman-Wright, K. (2008). The genome landscape of ERalpha- and ERbeta-binding DNA regions. *Proc. Natl. Acad. Sci. USA* **105**, 2604-2609.
- Lomvardas, S., Barnea, G., Pisapia, D. J., Mendelsohn, M., Kirkland, J. and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. *Cell* **126**, 403-413.
- Majumder, P., Gomez, J. A., Chadwick, B. P. and Boss, J. M. (2008). The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* **205**, 785-798.
- Martens, J. H., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P. and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24**, 800-812.
- Martin, S. and Pombo, A. (2003). Transcription factories: quantitative studies of nanostructures in the mammalian nucleus. *Chromosome Res.* **11**, 461-470.
- Miele, A. and Dekker, J. (2008). Long-range chromosomal interactions and gene regulation. *Mol. Biosyst.* **4**, 1046-1057.
- Miele, A., Bystricky, K. and Dekker, J. (2009). Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. *PLoS Genet.* **5**, e1000478.
- Misteli, T. (2001). The concept of self-organization in cellular architecture. *J. Cell Biol.* **155**, 181-185.
- Misteli, T. (2007). Beyond the sequence: cellular organization of genome function. *Cell* **128**, 787-800.
- Mitchell, J. A. and Fraser, P. (2008). Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* **22**, 20-25.
- Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., Goyenechea, B., Mitchell, J. A., Lopes, S., Reik, W. et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* **36**, 1065-1071.
- Pombo, A., Jackson, D. A., Hollinshead, M., Wang, Z., Roeder, R. G. and Cook, P. R. (1999). Regional specialization in human nuclei: visualization of discrete sites of transcription by RNA polymerase III. *EMBO J.* **18**, 2241-2253.
- Rando, O. J. and Chang, H. Y. (2009). Genome-wide views of chromatin structure. *Annu. Rev. Biochem.* **78**, 245-271.
- Reece-Hoyes, J. S., Deplancke, B., Shingles, J., Grove, C. A., Hope, I. A. and Walhout, A. J. (2005). A compendium of *Caenorhabditis elegans* regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol.* **6**, R110.
- Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E. et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306-2309.
- Sabo, P. J., Kuehn, M. S., Thurman, R., Johnson, B. E., Johnson, E. M., Cao, H., Yu, M., Rosenzweig, E., Goldy, J., Haydock, A. et al. (2006). Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. *Nat. Methods* **3**, 511-518.
- Sandhu, K. S., Shi, C., Sjölander, M., Zhao, Z., Göndör, A., Liu, L., Tiwari, V. K., Guibert, S., Emilsson, L., Imreh, M. P. et al. (2009). Nonallelic transvection of multiple imprinted loci is organized by the H19 imprinting control region during germline development. *Genes Dev.* **23**, 2598-2603.
- Schoenfelder, S., Sexton, T., Chakalova, L., Cope, N. F., Horton, A., Andrews, S., Kurukuti, S., Mitchell, J. A., Umlauf, D., Dimitrova, D. S. et al. (2010). Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet.* **42**, 53-61.
- Simonis, M., Klous, P., Splinter, E., Moshkin, Y., Willemsen, R., de Wit, E., van Steensel, B. and de Laat, W. (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat. Genet.* **38**, 1348-1354.
- Simonis, M., Kooren, J. and de Laat, W. (2007). An evaluation of 3C-based methods to capture DNA interactions. *Nat. Methods* **4**, 895-901.
- Solovei, I., Kreysing, M., Lantôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J. and Joffe, B. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* **137**, 356-368.
- Spellman, P. T. and Rubin, G. M. (2002). Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1**, 5.
- Spilianakis, C. G. and Flavell, R. A. (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat. Immunol.* **5**, 1017-1027.
- Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R. and Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature* **435**, 637-645.
- Spitz, F., Gonzalez, F. and Duboule, D. (2003). A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* **113**, 405-417.
- Sprout, D., Gilbert, N. and Bickmore, W. A. (2005). The role of chromatin structure in regulating the expression of clustered genes. *Nat. Rev. Genet.* **6**, 775-781.
- Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. *Exp. Hematol.* **33**, 256-271.
- Sutherland, H. and Bickmore, W. A. (2009). Transcription factories: gene expression in unions? *Nat. Rev. Genet.* **10**, 457-466.
- Thurman, R. E., Day, N., Noble, W. S. and Stamatoyannopoulos, J. A. (2006). Identification of higher-order functional domains in the human ENCODE regions. *Genome Res.* **17**, 917-927.
- Tiwari, V. K., Cope, L., McGarvey, K. M., Ohm, J. E. and Baylin, S. B. (2008a). A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res.* **18**, 1171-1179.
- Tiwari, V. K., McGarvey, K. M., Licchesi, J. D., Ohm, J. E., Herman, J. G., Schübeler, D. and Baylin, S. B. (2008b). PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* **6**, 2911-2927.
- Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. and de Laat, W. (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* **10**, 1453-1465.
- Vakoc, C. R., Letting, D. L., Gheldof, N., Sawado, T., Bender, M. A., Groudine, M., Weiss, M. J., Dekker, J. and Blobel, G. A. (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17**, 453-462.
- van Steensel, B. and Henikoff, S. (2000). Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* **18**, 424-428.
- Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. and Luscombe, N. M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252-263.
- Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. and Higgs, D. R. (2007). Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26**, 2041-2051.
- Versteeg, R., van Schaik, B. D., van Batenburg, M. F., Roos, M., Monajemi, R., Caron, H., Bussemaker, H. J. and van Kampen, A. H. (2003). The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes. *Genome Res.* **13**, 1998-2004.
- Vilar, J. M. and Saiz, L. (2005). DNA looping in gene regulation: from the assembly of macromolecular complexes to the control of transcriptional noise. *Curr. Opin. Genet. Dev.* **15**, 136-144.
- Walhout, A. J. (2006). Unraveling transcription regulatory networks by protein-DNA and protein-protein interaction mapping. *Genome Res.* **16**, 1445-1454.
- West, A. G. and Fraser, P. (2005). Remote control of gene transcription. *Hum. Mol. Genet.* **14**, R101-R111.
- Xu, M. and Cook, P. R. (2008). Similar active genes cluster in specialized transcription factories. *J. Cell Biol.* **181**, 615-623.
- Yu, M., Riva, L., Xie, H., Schindler, Y., Moran, T. B., Cheng, Y., Yu, D., Hardison, R., Weiss, M. J., Orkin, S. H. et al. (2009). Insights into GATA-1-mediated gene activation versus repression via genome-wide chromatin occupancy analysis. *Mol. Cell* **36**, 682-695.
- Zhao, R., Bodnar, M. S. and Spector, D. L. (2009). Nuclear neighborhoods and gene expression. *Curr. Opin. Genet. Dev.* **19**, 172-179.
- Zhao, Z., Tavossidana, G., Sjölander, M., Göndör, A., Mariano, P., Wang, S., Kanduri, C., Lezcano, M., Sandhu, K. S., Singh, U. et al. (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341-1347.
- Zhimulev, I. F., Belyaeva, E. S., Semeshin, V. F., Koryakov, D. E., Demakov, S. A., Demakova, O. V., Pokholkova, G. V. and Andreyeva, E. N. (2004). Polytene chromosomes: 70 years of genetic research. *Int. Rev. Cytol.* **241**, 203-275.
- Zhou, G. L., Xin, L., Song, W., Di, L. J., Liu, G., Wu, X. S., Liu, D. P. and Liang, C. C. (2006). Active chromatin hub of the mouse alpha-globin locus forms in a transcription factory of clustered housekeeping genes. *Mol. Cell Biol.* **26**, 5096-5105.