

3C-based methods to detect long-range chromatin interactions

Gang WEI (✉), Keji ZHAO (✉)

Laboratory of Molecular Immunology, National Heart, Lung and Blood Institute, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA

© Higher Education Press and Springer-Verlag Berlin Heidelberg 2011

Abstract Transcriptional regulatory regions are often located several thousand bases from the gene that they control. To function, the chromatin strand forms loops to juxtapose distal regions with the promoter. These long-range chromatin interactions have profound influences on the regulation of gene expression and mapping these interactions is currently a subject of intensive investigation. Chromosome conformation capture (3C) technology and its derivatives have been widely used to detect chromatin interactions and greatly contributed to understanding of the relationship between genome organization and genome function. Here we review these 3C-based methods for the study of long-range chromatin interactions and recent exciting findings obtained by using these technologies.

Keywords chromatin interactions, 3C, gene regulation, next-generation sequencing

Introduction

Sequencing a genome is only the first step to understand how it directs gene expression and function. Annotation of functional regulatory elements including promoters, enhancers, silencers, and insulators in the human genome and in other model organisms is a more challenging goal (ENCODE-consortium, 2004). Taking advantage of next-generation DNA sequencing technology, several genome-wide approaches have been developed to detect distribution of histone modifications and chromatin binding proteins (ChIP-Seq; Barski et al., 2007; Johnson et al., 2007), nucleosome position and occupancy (MNase-Seq; Schones et al., 2008), and nuclease hypersensitive sites (DNase-Seq; Boyle et al., 2008). Application of these techniques has provided an unprecedented understanding of the human epigenomes (Wang et al., 2009). Functional analysis of the epigenomic data has indicated that functional regulatory elements are associated with various histone methylation. Characterization of epigenomic data has provided a catalog of potential regulatory elements in the genomes. Although these studies have yielded important information of linear maps of genes and regulatory elements within the genome, there is a lack of information linking regulatory elements to their target genes. In addition, even though many genomic regions are

bound by transcription factors, only a fraction of these sites appear to be functional. Therefore, novel methods are needed to identify functional relationships between regulatory elements and target genes.

Emerging evidence suggests that the spatial organization of chromatin within the nucleus plays an important role in regulating genome functions such as transcription, DNA replication and DNA repair (Lanctôt et al., 2007; Misteli, 2007; Göndör and Ohlsson, 2009). Regulatory elements can act over large genomic distances to modulate genes expression by the formation of chromatin loops that physically link the element with its target genes. For example, looped chromatin conformation has been detected between the globin genes and the locus control region (LCR), which is located 35 kbp away from the beta globin promoter. This interaction is specific to erythroid cells where the gene is active, suggesting the long-range chromatin interaction is involved in transcriptional regulation (Carter et al., 2002). Mapping these long-range looped chromatin interactions will provide useful information on how regulatory sequences communicate with target genes.

Chromosome conformation capture (3C) is the preferred method to study chromatin interactions (Dekker et al., 2002) and has greatly contributed to our understanding of chromatin organization. 3C determines chromatin conformation by analyzing interaction frequencies between selected genomic sites. Many alterations of the 3C principle have been developed for large-scale applications including circular chromosome conformation capture (4C), chromosome conformation capture carbon copy (5C), chromatin interaction

Received September 30, 2010; accepted November 8, 2010

Correspondence: ^aGang WEI; ^bKeji ZHAO

E-mail: ^aweig2@mail.nih.gov; ^bzhaok@nhlbi.nih.gov

analysis by paired-end tag sequencing (ChIA-PET) and Hi-C (schematic representation in Fig. 1) (Dostie et al., 2006; Simonis et al., 2006; Fullwood et al., 2009; Lieberman-Aiden et al., 2009). All of these methods generated complementary evidence that the formation of chromatin loops allows interaction between genes and regulatory elements and led to great advances in our understanding of chromatin and genome organization. In this review, we will provide a brief summary of 3C-based techniques and the most interesting results obtained from application of these techniques.

3C technique

The 3C technology is the principle method for detecting long-range chromatin interactions at high-resolution (Dekker et al., 2002). The strategy of 3C and its derivatives is based on

the ‘proximity ligation’ concept, i.e. interacting chromatin segments are in close proximity and have greater chance to be ligated with each other. Briefly, formaldehyde is used to cross-link chromatin in living cells, keeping interacting chromatin segments close to each other through protein complexes that are bound to these segments. After cross-linking, chromatin is digested with a restriction enzyme and ligated under conditions that favor intramolecular ligation. After reversal of the cross-links, the DNA is purified and probed for long-range chromatin interactions with an anchor site using quantitative or semiquantitative PCR assays. The interaction frequency of two genomic fragments is measured by the relative abundance of the PCR product that is specific for the ligation product of these two fragments. It is important to note that the analysis of 3C data requires appropriate positive and negative controls.

Since the development of a standard protocol used to study

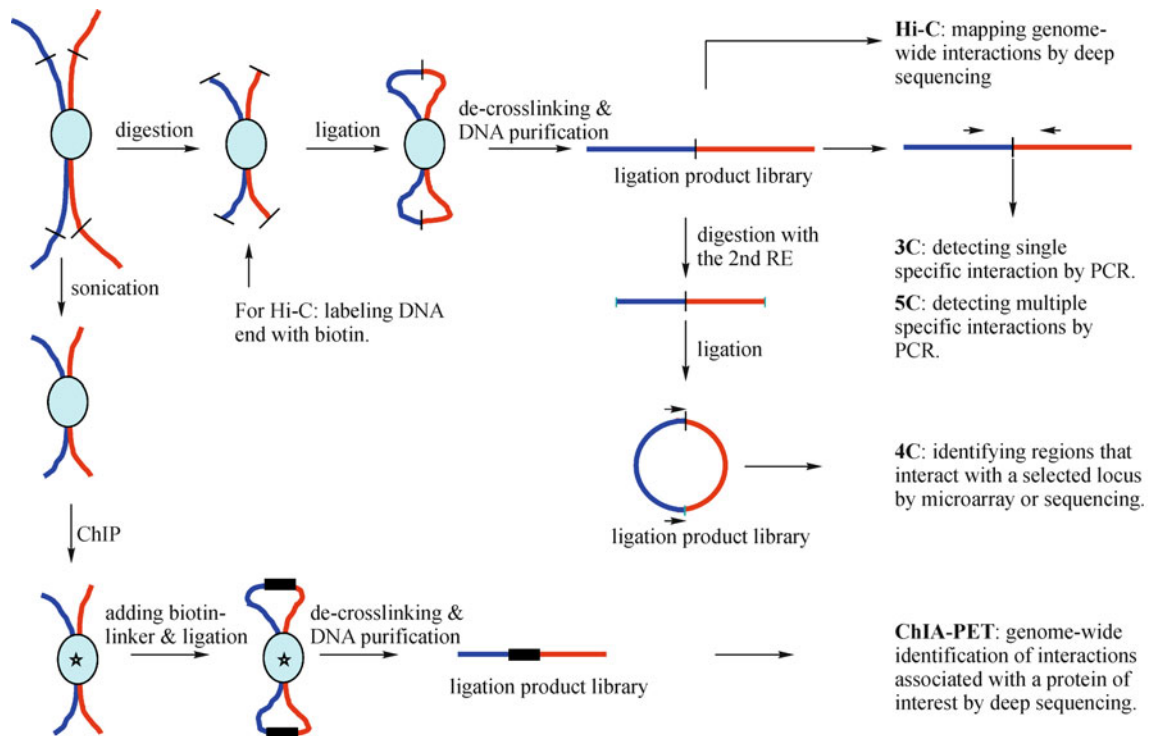


Figure 1 Schematic representation of the 3C-based methods. First, cells are fixed by formaldehyde treatment and interacted chromatin fragments that bind to the same protein complex are in close proximity. After digestion with a certain kind of restriction enzyme, chromatin fragments are re-ligated under diluted condition so that most of ligation reactions occur intramolecularly. The ligation products are subsequently purified, generating the conventional 3C-library. For 5C method, the conventional 3C library are denatured and annealed with a set of 5C primers so that multiple regions can be amplified with universal primers simultaneously. For 4C method, the conventional 3C library is further digestion with a second restriction enzyme and re-ligated intramolecularly, resulting in the circular ligation products, in which unknown regions that interacted with a pre-selected locus can be amplified by inverse PCR and identified by subsequent microarray analysis or deep sequencing. For Hi-C method, the digested chromatin fragments are filled in biotinylated dNTP before ligation so that the ligation products can be enriched by streptavidin beads for deep sequencing. In the case of ChIA-PET, the crosslinked chromatin is first fragmented by sonication instead of restriction enzyme digestion and chromatin immunoprecipitation is carried out to enrich the chromatin fragments that are associated with a specific protein. Biotin labeled half-linkers is ligated with the chromatin fragments and second ligation reactions occur intramolecularly through the cohesive ends of half linkers. The ligation products are subsequently digested by *MmeI* and the DNA fragments containing linkers are enriched by streptavidin beads selection for deep sequencing.

the spatial organization of yeast chromosome III (Dekker et al., 2002), the 3C technology has been successfully applied to analyze long-range chromatin interactions in several genomic loci in higher eukaryotic organisms. For example, long-range chromatin interactions between promoters and regulatory elements have been detected in the β -globin locus (Tolhuis et al., 2002; Palstra et al., 2003; Vakoc et al., 2005), the T-helper 2 cytokine locus (Spilianakis and Flavell, 2004), and the *Igf2/H19* imprinted locus (Murrell et al., 2004). In addition, *trans*-interactions between functional regulatory elements on different chromosomes have also been identified with the 3C method (Spilianakis et al., 2005; Xu et al., 2006). These studies confirm that chromatin looping is a general mechanism for distal elements to regulate gene expression.

5C technique

Although perfectly suited to probe hypothesized interactions, the conventional 3C method requires specific PCR primers and is not designed for unbiased or large-scale detection of genomic interactions. To perform high-throughput and comprehensive analysis of interaction networks within a large genomic region, the 5C technology was developed to detect long-range interactions between thousands of loci within a specific region (Dostie et al., 2006; Dostie and Dekker, 2006; Lajoie et al., 2009). In the 5C method, conventional 3C library is first generated which contains the ligation products of all chromatin interactions in the nucleus. Then the 3C library is denatured and hybridized to a mixture of 5C forward and reverse primers that are designed to only anneal across ligated junctions of the newly formed ligation products in the 3C library. After annealing to the 3C templates, the 5C primers are ligated with *Taq* ligase, which is followed by PCR amplification using universal primers that are annealed to the common tails of the 5C primers. Therefore, only the sequences around ligated junctions will be amplified and the resulting product is a 5C library. The relative abundance of a 5C product in the library reflects the frequency of interactions between genomic loci, which can be analyzed by either deep sequencing or microarray analysis.

The 5C method has been applied to analyze a 400-kb region containing the beta-globin locus and a 100-kb conserved gene desert region (Dostie et al., 2006). In this study, 78 5C primers were used to generate the 5C library and about 855 chromatin interactions were detected at various frequencies by microarray and quantitative sequencing. Theoretically, millions of chromatin interactions can be detected in a single assay with thousands of 5C primers.

Although 5C is a powerful tool to study chromatin interaction within a region at a relatively large scale, it requires very careful design of 5C primers in order to reduce amplification bias. In addition, 5C is not able to detect interactions with another unknown region beyond the region or even located on a different chromosome. Therefore, similar

with 3C, 5C is not an unbiased screen for genomic regions that interact with a specific DNA fragment.

4C technique

The 4C technique is designed to identify chromatin interactions between a given DNA fragment and the entire genome without any prior knowledge of the genomic regions involved. There are several different versions of 4C methods, but all of them depend on the formation of circular DNA molecules after the proximity ligation step, which is followed by the inverse PCR, using two bait-specific primers to amplify the circularized ligation products. Then the amplified products can be hybridized to probes on microarrays or deep-sequenced. The frequently detected genomic loci are identified as interacting regions of the fragment of interest.

The 4C method was first used to compare genomic regions that interact with the β -globin locus in an active or inactive conformation (Simonis et al., 2006). In this study, many long-range intrachromosomal and interchromosomal interactions were identified. Strikingly, the active β -globin locus in erythroid cells contacted a completely different set of loci compared with the inactive locus in brain cells: active β -globin gene prefers interacting with transcriptionally active regions whereas inactive β -globin gene tends to contact transcriptionally silent loci. This finding indicated that a switch of genomic environments is directly related to the transcriptional activities of the same locus in two different tissues. Similarly, other groups have also successfully applied 4C to identify regions interacting with the H19 imprinting control region (Zhao et al., 2006) and HoxB1 locus (Würtele and Chartrand, 2006).

ChIP-3C

Many proteins such as GATA1, Ezh2, CTCF and Brg1 are involved in the formation of chromatin loops and the depletion of these proteins could disrupt the chromatin interactions that they mediate (Vakoc et al., 2005; Tiwari et al., 2008; Splinter et al., 2006; Kim et al., 2009). Unlike 3C, 4C and 5C, which are used to detect general chromatin interactions, ChIP-3C (also called ChIP-loop) has been developed to identify chromatin interactions that are associated with specific proteins by the combination of 3C and Chromatin immunoprecipitation. In the ChIP-3C assay, the long-range interactions mediated by a specific protein are enriched by immunoprecipitation before the proximity ligation step in the conventional 3C protocol. Using the ChIP-3C method, chromatin interactions mediated by several proteins including chromatin loop at Dlx5-Dlx6 locus mediated by MeCP2, interactions between ER α binding sites by ER α and chromatin interactions at the Th2 cytokine locus by SATB1 have been examined (Horike et al., 2005;

Carroll et al., 2005; Cai et al., 2006). These studies provide great insights into how important proteins regulate gene expression by controlling chromatin interactions. However, the analysis of ChIP-3C result is more complicated than conventional 3C because the abundance of specific ligation products is also affected by the relative enrichment of chromatin fragments by the ChIP step. Furthermore, ChIP-3C, like the other 3C variations mentioned focuses on a single protein and is not an unbiased genome-wide application.

ChIA-PET

ChIP-3C can be extended beyond a single region to the entire genome with next generation sequencing. The 3C ligation products can be analyzed by sequencing both ends of a DNA fragment (paired-end ditag, PET). Detection of two sequences, which normally are located far away from each other in the genome, within one DNA fragment indicates a long-range interaction. ChIA-PET, a combination of 3C, ChIP and PET, provides the power of detecting chromatin interactions mediated by a specific protein, as demonstrated for ER α , in the whole genome (Fullwood et al., 2009). In this technique, crosslinked chromatin is fragmented by sonication instead of restriction enzyme digestion and chromatin fragments bound by specific proteins are enriched by ChIP. A biotinylated DNA half-linker is ligated to the end of enriched chromatin fragments and then two chromatin fragments with half-linker are proximately ligated via the linker bridge. The ligation products are digested with *MmeI* as the linker sequence contains *MmeI* recognition site. The resulting biotinylated tag-linker-tag fragments are enriched with streptavidin beads and ligated to specific adaptors, followed by PCR amplification and next-generation pair-end sequencing.

A study by Fullwood et al. (2009) provided the first genome-wide map of chromatin interactome bound by ER α at base-pair resolution. The authors identified 1451 intrachromosomal and 15 interchromosomal interactions. Many distal binding sites of ER α are linked to specific estrogen-responsive genes over a long genomic distance and may contribute critically to transcriptional regulation of these genes. In addition, many loci containing ER α binding sites contact each other extensively, suggesting these loci are looped into a common protein complex and their transcription activities may be coordinately regulated. Therefore, this method could provide important functional information regarding protein-mediated chromatin interactions.

Hi-C

Unlike ChIA-PET, which focuses on the interactions mediated by a specific protein, the recently developed Hi-C technique is used to map chromatin interactions in the entire

genome, although at relatively low resolution (1 megabases) (Lieberman-Aiden et al., 2009). The protocol of Hi-C is similar to conventional 3C with the modification that biotinylated nucleotides are incorporated into the DNA ends before the proximity ligation step. DNA is then sheared by sonication and the ligation junctions are enriched by binding to streptavidin beads. The resulting Hi-C library is analyzed by high throughput sequencing, which leads to unbiased identification of global chromatin interactions.

Several features of genomic organization were revealed by the Hi-C data in this study. First, the presence of chromosome territories is confirmed by higher contact probability of genomic loci within the same chromosome than those in separate chromosomes. Second, correlation analysis shows two general compartments in the human genome: one compartment contains open and active chromatin and the other contains close and inactive chromatin regions. Lastly, data analysis indicates that chromatin resembles a fractal globule, a knot-free, tightly packed conformation rather than a compact, densely knotted structure.

Another group has developed a 4C-based approach to generate a three-dimensional map of the yeast genome at kilobase resolution, revealing the complexity of genome organization even in the simple organism (Duan et al., 2010). Similar with the study of the human genome, the resolution of the map is constrained by the cost of deep sequencing. As the cost of sequencing continually decreases, it will be expected to get this kind of map at much higher resolution.

Future perspectives

One big challenge in characterizing human epigenomes is to map chromatin interactions in three dimensions and understand how these interactions influence transcription and other functions of the genome. Over the last several years, 3C-based technologies have pushed the field forward and contributed enormously to our understanding of the relationship between chromatin topology and genome function. In addition, with the advent of the next generation sequencing technologies, ChIA-PET and Hi-C methods have been recently developed to allow genome-wide mapping of interactions mediated by a specific protein or unbiased chromatin interactions, respectively. Eventually, high-resolution maps of chromatin interactome will be available in future, which could link every regulatory element with its target gene. It needs to be mentioned that all these 3C-based methods detect chromatin interactions in a large population of cells and additional approaches such as FISH (fluorescence *in situ* hybridization) and other real time imaging methods are required to confirm the interactions in single cells or even in living cells.

Together, these powerful tools for mapping chromatin interactions can be used to address many open questions in the field. For example, how does genome configuration

change during cell cycle or developmental processes? How does a specific transcription factor or chromatin binding protein regulate the genome configuration? Are chromatin interactions dynamically regulated and do they contribute to epigenetic and transcriptional regulation? Answering these questions will help to understand the relationship between chromatin interactome and genome function. Further studies integrating 3C-based methods with other genetic, biochemical approaches to address questions will provide new insight into gene regulation in three dimensions.

Acknowledgements

We thank Dr. Daniel Northrup for critical reading of the manuscript. Research in the authors' laboratory is supported by the Intramural Research Program of the National Heart, Lung and Blood Institute, National Institutes of Health, USA.

Reference

- Barski A, Cuddapah S, Cui K R, Roh T Y, Schones D E, Wang Z B, Wei G, Chepelev I, Zhao K (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4): 823–837
- Boyle A P, Davis S, Shulha H P, Meltzer P, Margulies E H, Weng Z, Furey T S, Crawford G E (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell*, 132(2): 311–322
- Cai S T, Lee C C, Kohwi-Shigematsu T (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. *Nat Genet*, 38(11): 1278–1288
- Carroll J S, Liu X S, Brodsky A S, Li W, Meyer C A, Szary A J, Eeckhoutte J, Shao W L, Hestermann E V, Geistlinger T R, Fox E A, Silver P A, Brown M (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell*, 122(1): 33–43
- Carter D, Chakalova L, Osborne C S, Dai Y F, Fraser P (2002). Long-range chromatin regulatory interactions *in vivo*. *Nat Genet*, 32(4): 623–626
- Dekker J, Rippe K, Dekker M, Kleckner N (2002). Capturing chromosome conformation. *Science*, 295(5558): 1306–1311
- Dostie J, Dekker J (2006). 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 D, Krumm A, Lamb J, Nusbaum C, Green R D, Dekker J (2006). Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res*, 16(10): 1299–1309
- Duan Z, Andronescu M, Schutz K, McIlwain S, Kim Y J, Lee C, Shendure J, Fields S, Blau C A, Noble W S (2010). A three-dimensional model of the yeast genome. *Nature*, 465(7296): 363–367
- ENCODE project consortium (2004). The ENCODE (ENCyclopedia of DNA Elements) Project. *Science*, 306(5696): 636–640
- 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, Chew E G, Huang P Y, Welboren W J, Han Y, Ooi H S, Ariyaratne P N, Vega V B, Luo Y, Tan P Y, Choy P Y, Wansa K D, Zhao B, Lim K S, Leow S C, Yow J S, Joseph R, Li H, Desai K V, Thomsen J S, Lee Y K, Karuturi R K, Herve T, Bourque G, Stunnenberg H G, Ruan X, Cacheux-Rataboul V, Sung W K, Liu E T, Wei C L, Cheung E, Ruan Y (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature*, 462(7269): 58–64
- Göndör A, Ohlsson R (2009). Chromosome crosstalk in three dimensions. *Nature*, 461(7261): 212–217
- Horike S, Cai S T, Miyano M, Cheng J F, Kohwi-Shigematsu T (2005). Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat Genet*, 37(1): 31–40
- Johnson D S, Mortazavi A, Myers R M, Wold B (2007). Genome-wide mapping of *in vivo* protein-DNA interactions. *Science*, 316(5830): 1497–1502
- Kim S I, Bresnick E H, Bultman S J (2009). BRG1 directly regulates nucleosome structure and chromatin looping of the alpha globin locus to activate transcription. *Nucleic Acids Res*, 37(18): 6019–6027
- Lajoie B R, van Berkum N L, Sanyal A, Dekker J (2009). My5C: web tools for chromosome conformation capture studies. *Nat Methods*, 6(10): 690–691
- Lancôt C, Cheutin T, Cremer M, Cavalli G, Cremer T (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet*, 8(2): 104–115
- Lieberman-Aiden E, van Berkum N L, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie B R, Sabo P J, Dorschner M O, Sandstrom R, Bernstein B, Bender M A, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny L A, Lander E S, Dekker J (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science*, 326(5950): 289–293
- Misteli T (2007). Beyond the sequence: cellular organization of genome function. *Cell*, 128(4): 787–800
- Murrell A, Heeson S, Reik W (2004). Interaction between differentially methylated regions partitions the imprinted genes *Igf2* and *H19* into parent-specific chromatin loops. *Nat Genet*, 36(8): 889–893
- Palstra R J, Tolhuis B, Splinter E, Nijmeijer R, Grosveld F, de Laat W (2003). The beta-globin nuclear compartment in development and erythroid differentiation. *Nat Genet*, 35(2): 190–194
- Schones D E, Cui K, Cuddapah S, Roh T Y, Barski A, Wang Z, Wei G, Zhao K (2008). Dynamic regulation of nucleosome positioning in the human genome. *Cell*, 132(5): 887–898
- Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006). Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet*, 38(11): 1348–1354
- Spilianakis C G, Flavell R A (2004). Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol*, 5(10): 1017–1027
- Spilianakis C G, Lalioti M D, Town T, Lee G R, Flavell R A (2005). Interchromosomal associations between alternatively expressed loci. *Nature*, 435(7042): 637–645
- Splinter E, Heath H, Kooren J, Palstra R J, Klous P, Grosveld F, Galjart N, de Laat W (2006). CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev*, 20(17): 2349–2354
- Tiwari V K, Cope L, McGarvey K M, Ohm J E, Baylin S B (2008). A

- novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations. *Genome Res*, 18(7): 1171–1179
- Tolhuis B, Palstra R J, Splinter E, Grosveld F, de Laat W (2002). Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell*, 10(6): 1453–1465
- Vakoc C R, Letting D L, Gheldof N, Sawado T, Bender M A, Groudine M, Weiss M J, Dekker J, Blobel G A (2005). Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell*, 17(3): 453–462
- Wang Z, Schones D E, Zhao K (2009). Characterization of human epigenomes. *Curr Opin Genet Dev*, 19(2): 127–134
- Würtele H, Chartrand P (2006). 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
- Xu N, Tsai C L, Lee J T (2006). Transient homologous chromosome pairing marks the onset of X inactivation. *Science*, 311(5764): 1149–1152
- Zhao Z, Tavossidana G, Sjölander M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu K S, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R (2006). Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet*, 38(11): 1341–1347