npg

RESEARCH HIGHLIGHT

DNA interaction networks: an information highway for regulated gene expression in the 3-dimentional space of the nucleus

Dong Wang[1], Xiang-Dong Fu[1]

[1]Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0651, USA

In this genomic era, the research landscape has been taking a dramatic turn in discovering new biological principles that have been unapproachable from traditional studies on the basis of individual genes and this process is accelerating because of technological breakthroughs, largely fueled by the next generation sequencing. Many years of research have established a dogmatic view on transcription that gene promoters drive transcriptional initiation, which is subject to modulation by distal enhancers [1]. Most biochemically dissected enhancers are localized in the vicinity of gene promoters and are rarely beyond gene boundaries. This traditional view of gene units is now challenged by genome-wide location analysis in recent years, which reveals that many DNA binding transcription factors bind to genomic loci that can be far away from their regulated genes. For example, the activated estrogen receptor α (ERα) was found to largely bind to intergenic regions in the human genome, rather than near gene promoters responsive to estrogen stimulation [2]. This and many other similar studies raise a general question with respect to whether some of those binding events are fortuitous or functional, and if functional, how these enhancers find their target promoters to confer signal-induced regulation of gene expression.

In order to avoid interference of transcriptional activities of one transcription unit with another, there also exist DNA elements called insulators bound by the CTCF protein that prevent enhancers from influencing their multiple neighboring genes, but understanding of how insulators work is an evolving process. The insulator complex was initially thought to serve as a roadblock to tracking of transcription complexes assembled on enhancers to target gene promoters. While this tracking model is consistent with many lines of experimental evidence from studying individual gene units, it is difficult to envision how a transcription complex is able to travel a long distance, especially when across one or multiple other active transcription units, to reach its final destination. The DNA looping model was thus proposed to explain long distance enhancer-promoter interactions, which are likely instrumented by interactions of protein complexes formed on both enhancers and promoters [3]. In this regard, an insulator may function through enhancing some specific long-distance interactions while suppressing others to facilitate specific partnership between enhancers and promoters, which is consistent with recent genomic studies [4].

Last, but not least, genome-wide association scanning (GWAS) has revealed that genomic landmarks (often in the form of single nucleotide polymorphisms (SNPs)) genetically linked to specific human diseases are frequently located in gene desert regions in the genome, the latest example being a locus associated with coronary heart disease in human chromosome 9 [5]. These findings challenge the traditional wisdom of gene-central disease mechanisms and raise the possibility that many gene desert regions may provide critical regulatory functions through long-distance DNA-DNA interactions within chromosomes or even between chromosomes.

Clearly, while DNA molecules are linear, two functional DNA elements in a string of DNA may be a lot closer than what their physical distance implies. Rather than random collision, they may specifically interact with one another aided by the protein machineries assembled on them. To detect such long-distance interactions, a key technology was intervened in 2002, which is referred to the Chromosome Conformation Capture (3C) assay [6]. The 3C assay makes the assumption that, after restriction digestion, DNA

Correspondence: Xiang-Dong Fu
Tel: +858-534-4937; Fax: +858-822-6920
E-mail: xdfu@ucsd.edu

segments that are tethered by common protein complexes have greater kinetic advantages in the DNA ligation reaction under dilute conditions than those freely diffusing in solution or anchored in different complexes, and once being ligated, the shuffled DNA products can be detected by PCR (Figure 1A). This proximity-based method was initially designed to interrogate specific candidate DNA-DNA interactions in question. While various improvements and extension of this 3C method have been made for more unbiased applications, it has been a dream to develop a truly unbiased approach to detect DNA-DNA interactions genome-wide.

This has now been accomplished by two teams, one led by Drs Cheung and Ruan in Singapore [7] and the other by Drs Lander and Dekker from the US [8]. The Singapore group developed the ChIA-PET technology to detect DNA-DNA interactions that are tethered by a common transcription factor (Figure 1B). The procedure starts with the initial steps in the traditional 3C

assay followed by a chromatin immunoprecipitation step to enrich for ERα-bound complexes. A biotinylated DNA linker that contains a site for a type II restriction enzyme (such as *Mme*I, which cuts DNA at a distance from its recognition site) is ligated to individual DNA ends, which are next linked in a second ligation reaction to result in the insertion of the dual linker between two DNA fragments tethered by common DNA-protein complexes. Upon restriction digestion to release the tag-linker-tag fragments and enrichment of such fragments by streptavidin affinity selection, specific adaptors are added to the fragments for PCR amplification and then for high throughput sequencing. In comparison, the Hi-C technology developed by the US group is simplified by incorporating biotinylated nucleotides into DNA ends prior to the ligation step in the standard 3C reaction, which allows affinity selection for shuffled DNA, adaptor addition, and high throughput sequencing (Figure 1C). Because of the omission of the anti-

body enrichment step, the Hi-C method permits genome-wide detection of long-distance DNA-DNA interactions, but without the information on specific interactions mediated by defined proteins, which should be amenable by incorporating the immunoprecipitation step as in ChIA-PET. Although both methods detected a large number of re-ligated DNA at the original cleavage sites, massive sequencing allowed detection of shuffled DNA, a testimony for the power of high throughput sequencing technologies that are transforming modern biological research.

What have we learnt from these power technological innovations? The Singapore group applied ChIA-PET to network interactions mediated by liganded estrogen receptor α in a breast cancer cell line, detecting a large number of intra-chromosomal interaction clusters that stand above the background. Although most of the intra-chromosomal interactions (86%) take place between loci separated by < 100 Kb, a significant fraction of the in-
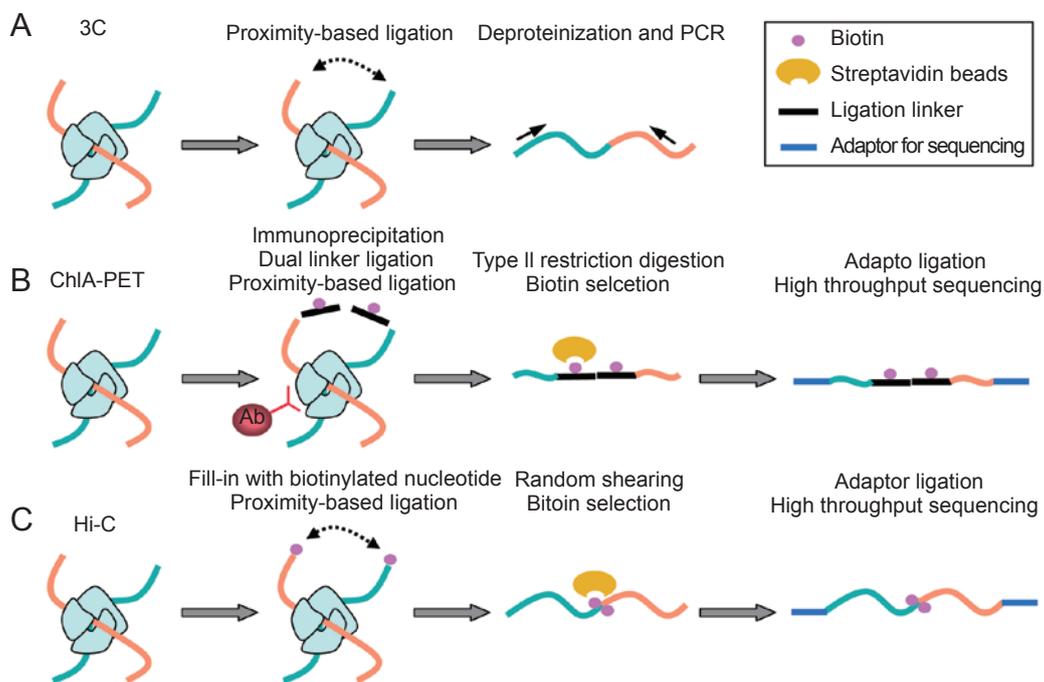

Figure 1 Illustration of the Chromosome Conformation Capture (3C) assay **(A)** and the newly developed ChIP-PET **(B)** and Hi-C **(C)** technologies for global 3C analysis.

teractions appear to occur between loci that are 100 Kb or more away from one another. A panel of these interactions was further confirmed by conventional 3C and FISH, some of which were also shown to be inducible by estrogen.

Interestingly, the authors detected two types of long-distance intra-chromosomal interactions. Besides duplex interactions, which reflect individual paired interactions between two specific genomic loci, they also observed many complex interactions that involve multiple ERα-bound DNA loci, suggesting that many DNA segments are looped into a common protein complex. These findings therefore suggest extensive DNA looping in the genome and collective interactions that may underlie coordinated regulation of gene expression from those interacting loci. The data linked the distal ERα binding sites to specific estrogen-responsive genes, and in many those specific pairs, ERα and Pol II appear to exhibit reciprocal binding strength (more ERα at the distal than the promoter-proximal binding sites and the converse is true for Pol II), which is in line with the looping model for enhancer-promoter communications during transcription initiation. More importantly, the interaction map allowed the authors to assign genes that are intimately associated with DNA-DNA interactions (called anchored genes) versus genes resident in looped regions (called looped genes). Microarray analysis demonstrated that anchored genes are more likely up-regulated by liganded ERα than the genes in the looped regions. The dataset offers a rich source for individual investigators to look for key regulatory elements in their favorite gene models for detailed mechanistic studies.

The US group focused on using the Hi-C data to decipher some general principles underlying chromosomal organization in the 3-dimensional space of the nucleus, which are clearly non-random events as recently reviewed [9]. The authors mined the data with a range of statistical tools to calculate the contact probability of DNA loci as the function of genomic distance. This analysis first confirmed chromosome territories observed by numerous chromosome painting studies, because DNA loci within the chromosomes (intra-chromosome interactions) always show a great contact probability than those in separate chromosomes (inter-chromosomal interactions), therefore implying that DNA sequences within individual chromosomes occupy distinct territories in the nucleus. By calculating the contact probability within different regions of single chromosomes (currently at the resolution of 1 Mb), the authors also recognized two general patterns of chromosomal segments consisting of open versus closed (or packed) chromatin regions, which likely correspond to eu- and hetero-chromatins, respectively, as defined by biochemical and cytological studies. As expected, the regions associated with open chromatins are more closely linked than packed chromatins to a variety of genetic (DNase I hypersensitive sites) and epigenetic (histone modifications) marks associated with regulated gene expression.

Perhaps one of the most revealing aspects of the current study is the application of the Hi-C data to test models of DNA configuration within individual chromosome territories. When considered a chromosome as a form of polymer with various local and long-distance interactions, the so-called equilibrium globule model suggests a series of globular domains that are packed (after reaching the equilibrium) in a relatively random fashion in the nucleus. This model does not accommodate a mechanism to prevent the formation of knots, which would obviously obstruct the chromosome condensation/decongestion process during the cell cycle. The alternative fractal globule model emphasizes a self-organizing process of a chromosome consisting of a series of small globular domains, resulting in an unentangled, knot-free polymer. The fractal model appears to make more sense, which is supported by 3D distance measured in a recent FISH study [10]. This conclusion is now further substantiated by curve fitting of the Hi-C derived contact probability as a function of genomic distance.

In summary, the genome-wide analysis of long-distance DNA-DNA interactions reported by the two recent studies marks the beginning of a predictable new wave of investigation into the physical basis of regulated gene expression. The data revealed by ChIP-PET has already suggested many co-regulated events that are likely the driver for the formation of small globules proposed in the fractal globule model now backed by the Hi-C data. Applications of these genome-wide mapping technologies in diverse biological systems will allow assignment of specific intergenic binding sites bound by individual transcription factors or critical SNPs from GWAS studies to their target genes. CTCF has been implicated in certain long-distance interactions [11], which can now be extended genome-wide to elucidate its role in coordinated gene expression as well as gene partitioning into individual interacting chromosomal domains as part of its insulator function. One technical challenge is an exponential number of tag density needed for a linear increase in detecting DNA-DNA interactions, especially for those taking place between chromosomes that depend on activating signals [12, 13]. This may be addressed by coupling the unbiased, genome-wide mapping strategies with target-specific approaches, such as the ligation-based 5C technique [14], to derive quantitative information on specific intra- and inter-chromosomal interactions. Obviously, these genomic technologies, which give rise to results derived from molecular averaging in a large number of cells, also need to be coupled with single cell-based, direct imaging techniques to study the specificity and dynamics of DNA folding/unfolding and chromosome rearrangement in response to

signaling and to link specific molecular events to morphologically defined cellular structures, such as nuclear lamina [15] and nuclear speckles enriched with various transcription factors and the splicing machinery [12, 16]. With the current pace of technological innovation, it is virtually certain that we will witness an accelerated speed of discoveries in years ahead.

References

1 Blackwood EM, Kadonaga JT. Going the distance: a current view of enhancer action. *Science* 1998; **281**:60-63.

2 Carroll JS, Meyer CA, Song J, *et al*. Genome-wide analysis of estrogen receptor binding sites. *Nat Genet* 2006; **38**:1289-1297.

3 Bulger M, Groudine M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* 1999; **13**:2465-2477.

4 Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* 2009; **137**:1194-1211.

5 McPherson R, Pertsemlidis A, Kavaslar N, *et al*. A common allele on chromosome 9 associated with coronary heart disease. *Science* 2007; **316**:1488-1491.

6 Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002; **295**:1306-1311.

7 Fullwood MJ, Liu MH, Pan YF, *et al*. An oestrogen-receptor-α-bound human chromatin interactome. *Nature* 2009; **462**:58-64.

8 Lieberman-Aiden E, van Berkum NL, Williams L, *et al*. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009; **326**:289-293.

9 Gondor A, Ohlsson R. Chromosome crosstalk in three dimensions. *Nature* 2009; **461**:212-217.

10 Mateos-Langerak J, Bohn M, de Leeuw W, *et al*. Spatially confined folding of chromatin in the interphase nucleus. *Proc Natl Acad Sci USA* 2009; **106**:3812-3817.

11 Ling JQ, Li T, Hu JF, et al. CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* 2006; **312**:269-272.

12 Hu Q, Kwon YS, Nunez E, *et al*. Enhancing nuclear receptor-induced transcription requires nuclear motor and LSD1-dependent gene networking in interchromatin granules. *Proc Natl Acad Sci USA* 2008; **105**:19199-19204.

13 Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. *Nature* 2005; **435**:637-645.

14 Dostie J, Richmond TA, Arnaout RA, *et al*. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 2006; **16**:1299-1309.

15 Guelen L, Pagie L, Brasset E, *et al*. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 2008; **453**:948-951.

16 Brown JM, Green J, das Neves RP, *et al*. Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J Cell Biol* 2008; **182**:1083-1097.