Meet the neighbours: tools to dissect nuclear structure and function

Cameron S. Osborne, Philip A. Ewels* and Alice N.C. Young*

Advance Access publication date 21 January 2011

Abstract
The eukaryotic cell nucleus displays a high degree of spatial organization, with discrete functional subcompartments that provide microenvironments where specialized processes take place. Concordantly, the genome also adopts defined conformations that, in part, enable specific genomic regions to interface with these functional centers. Yet the roles of many subcompartments and the genomic regions that contact them have not been explored fully. More fundamentally, it is not entirely clear how genome organization impacts function, and vice versa. The past decade has witnessed the development of a new breed of methods that are capable of assessing the spatial organization of the genome. These stand to further our understanding of the relationship between genome structure and function, and potentially assign function to various nuclear subcompartments. Here, we review the principal techniques used for analyzing genomic interactions, the functional insights they have afforded and discuss the outlook for future advances in nuclear structure and function dynamics.

Keywords: genome organization; nuclear structure and function; chromosome conformation capture; next generation sequencing

Transcription of the genome is fundamentally regulated at three distinct yet inter-dependent levels. Firstly, trans-acting regulatory factors are attracted to specific DNA sequences to ensure genes are transcribed in the right cells, at the right time [1]. The second level of control is achieved through modification of both the DNA and the nucleosomes that package it into chromatin [2]. These epigenetic marks can act to recruit activating or repressive protein complexes, and modulate activity by effecting chromatin compaction. Finally, non-random organization of the genome within the nucleus appears to influence transcription [3]. Indeed, the nucleus is a highly structured and compartmentalized organelle. Many subcompartments have been identified by microscopy and biochemical analysis; some of these are relatively well characterized, while the functions of others remain elusive. It is likely that specific regions of the genome contact different subcompartments to carry out their individual roles.

Arrangement of the genome such that certain regions contact the various nuclear subcompartments is perhaps predictable, yet interestingly, it is a responsive, adaptable and cell-type specific process. At interphase, individual chromosome territories occupy discrete positions in the nucleus [4–6], although at least a certain degree of intermingling between neighboring territories is evident [7]. Each chromosome often adopts a preferred position within the nucleus, and displays specific preferences for neighboring chromosomes. Remarkably, this organization is cell-type specific [8], implying that chromosomal...
positioning may reflect or direct differential gene expression. Indeed, changes in the preferred neighboring of chromosomes occur as progenitor cells differentiate into mature cells [9]. The speed with which the whole chromosome repositioning takes place is astonishing; Bridger and colleagues have observed large-scale chromosome movements in fibroblast nuclei within 15 min of quiescence-inducing serum starvation [10]. Significantly, these movements require energy and actin/myosin polymerization, suggesting that genome reorganization is a directed process. The tight relationship between genome conformation and gene activity is further supported by a study that artificially and reversibly tethered a genomic region to the repressive nuclear periphery, resulting in transcriptional suppression [11, 12], although this is not always the case [13]. Therefore, it appears likely that genomic rearrangements can play a driving role in gene expression changes, rather than merely reflecting a passive consequence.

Over the past 30 years, the limelight has been fixed upon the contributions to genetic control by specific DNA elements, and subsequently chromatin structure modifications, yet the involvement of nuclear organization has remained enigmatic and in the shadows. The major advances have come mostly through microscopy-based analysis, which have been instrumental in the identification and characterization of nuclear subcompartments, and the non-random organization of the genome. However, these efforts have been hampered by both resolution limitations of light microscopy and difficulties in identifying the DNA elements with which the subcompartments interface.

An important technological breakthrough occurred in 2002, when new methodologies were developed to study genomic spatial arrangements of the genome. Firstly, Dekker and colleagues seized upon the ‘Nuclear Ligation Assay’ from the Seyfred laboratory [14, 15] to develop the chromosome conformation capture method (3C) (Figure 1A), a DNA ligation-based proximity assay that was used to examine the structural organization of yeast chromosomes in the nucleus [16]. This provided the first description of a genomic conformation in situ.

3C was rapidly adapted to measure long-range interactions in mammalian cells. Along with a separate, newly developed method called RNA FISH TRAP, which identifies the genomic spatial arrangements at the site of a specific transcript [17], 3C was used to capture the close, physical association that occurs between the β-globin gene promoter and its locus control region (LCR) located 50 kb upstream [18]. This result was significant because it provided strong evidence that distal regulatory elements act upon their target promoters through direct contact, while looping out intervening sequences. This highlighted that gene regulation occurs in three dimensions. Whereas RNA FISH TRAP is technically rather challenging, the relative simplicity of 3C has led to its widespread adoption in studies of long-range interactions at numerous gene loci. The linear distance separating specific interacting regions can be staggering; some cis-interacting partners are located over a megabase apart [19]. Even more remarkable, 3C experiments have suggested specific interactions can occur between loci on separate chromosomes [20, 21]. However, not all interactions detected by 3C necessarily indicate a specific and functional contact between loci. Detected interactions may also occur through co-associations at shared nuclear subcompartments, such as actively transcribed genes at transcription factories [22, 23].

A frustrating limitation of 3C is that it relies upon detection of interacting loci using specific polymerase chain reaction (PCR) primers, yet within a 3C library, a full gamut of genome-wide interactions is present. This precludes the detection of unexpected, yet important interactions. Several groups have addressed this limitation in an effort to identify all the loci that interact with a specific sequence of interest. In essence, two general approaches have emerged, each with varying degrees of success in detecting unpredicted interactions.

The first approach can be categorized generally as circularized 3C (Figure 1B). Here, DNA circles that contain the ‘bait’ sequence and an interacting partner sequences are studied. The DNA circles are either formed naturally through the 3C procedure [24, 25], or generated during subsequent steps by a second restriction enzyme digestion and ligation [26, 27]. Finally, the interacting partner sequences are amplified by inverse PCR and identified by either microarray or sequencing. Württele and Chartrand used their ‘open-ended 3C’ assay to investigate the spatial environment of the HoxB1 gene during its induction [27]. Zhao et al. [25] used a technique termed 4C (circular 3C or 3C-on-chip) to uncover extensive networks of epigenetically regulated chromosomal interactions, both in cis and in trans. Work by Simonis et al. [26] using 4C also gave insight into
Figure 1: (A) Traditional chromosome conformation capture (3C). DNA sequences co-localising within a region of the nucleus are fixed using formaldehyde and cut using a restriction endonuclease such as Bgl II or Hind III. Cut DNA ends are joined in dilute conditions, favoring ligation events within a fixed complex. Cross-links are reversed and DNA is purified. PCR and quantitative PCR can be used with specific primers to detect expected products in a quantitative manner. (B) Circularized 3C, as carried out by Zhao et al. and Lomvardas et al. [24, 25]. Würtele et al. and Simonis et al. use a second round of restriction digestion and ligation to reduce the DNA circle size [26, 27] (data not shown). DNA sequences are subjected to a prolonged ligation step to form circles. Nested PCR using primers within the known ‘bait’ sequence is performed, yielding a library with known end fragments containing
nuclear organization by showing that active and inactive chromatin domains form distinct interaction clusters, which is in keeping with the long-held view of active and silent chromatin being spatially segregated into euchromatin and heterochromatin, respectively. Interestingly, circularized 4C approaches detect relatively few long-range cis and trans interactions compared with other 4C methodologies, although it is uncertain why this is the case.

The second approach used to capture interactions of a specific locus can be classified as ‘adapter 3C’ methodologies (Figure 1C). Like some iterations of circular 3C, these techniques involve a second restriction enzyme digestion step to cut the DNA within the interaction partner sequence; however, they diverge when an adapter sequence is ligated to the sticky end. The library is amplified using primers that hybridize within the adapter and bait sequences. The method was first employed by Ling et al. [28] in their ‘associated chromosome trap’ (ACT) assay, which identified three potential interaction partners of an imprinting control region within the Igf2-H19 locus on mouse chromosome 7. Of the three interactions, one was intrachromosomal and known previously, while the remaining two were interchromosomal and novel. However, the assay failed to detect other regions known to interact with the Igf2-H19 locus [29, 30], suggesting that the list of targets identified by the screen was not comprehensive; again, the underlying reasons for this are unclear [31].

A related technique was developed by Schoenfelder et al. [32], called enhanced ChIP 4C (e4C), which incorporates two major modifications to the method used by Ling et al., namely immunoprecipitation and biotin enrichment. Immunoprecipitation has been employed before in 3C-derived methods; both the ChIP loop and combined 3C-ChIP cloning (or 6C) assays contain an immunoprecipitation step before or after ligation, respectively [33, 34]. The enriched ligation products are either detected by specific PCR primers, or cloned and sequenced. In the e4C studies, an antibody that recognizes the transcription-initiating form of RNA polymerase II was used to enrich for the transcribed regions of the genome, to focus on interactions that occur within transcription factories. Secondly, further enrichment is achieved by annealing a biotinylated, bait-specific primer and primer extending into the 3C-ligated interaction partner sequences. By isolating the biotinylated bait ligation fragments on streptavidin-coated magnetic beads, library complexity is greatly reduced, which greatly increases the sensitivity for detection of specific interactions involving a region of interest. Indeed, Schoenfelder et al. [32] found a higher incidence and richer assortment of long-range cis and trans interactions between genes than had previously been detected by other ‘4C’-like incarnations. Significantly, they found evidence that genes preferentially co-associate with distinct subsets of other genes, which appears to be in part a reflection of the shared trans-acting factors that regulate them, and highlights that transcription may act as a major influencing force on genome organization.

Efforts of late have moved toward capturing multiple interacting genomic regions concurrently to provide a more comprehensive view of genome organization. Dostie et al. [35] made the first step, developing 5C (chromosome conformation capture carbon copy) to map the interactions that occur within a given gene locus. This technique uses ligation-mediated amplification, where oligonucleotide primers that anneal immediately adjacent to 3C restriction sites are ligated together to generate an interaction library of fusion oligonucleotides that can be assayed by sequencing or microarray (Figure 1D). Dostie and colleagues applied 5C to the human β-globin gene locus, where they confirmed previously studied interactions and also identified new ones. While there will be practical and financial limitations to the degree of 5C multiplexing that is achievable, this method remains

Figure 1: Continued
unknown interacting sequences. (C) Adapter ligation 3C. A second restriction enzyme digestion step creates a sticky end near the ligation junction. An adapter sequence is ligated to this sticky end, enabling PCR amplification of products using primers within the ‘bait’ sequence and adapter. (D) 5C. Multiple ligation-mediated amplification primers containing adapter sequence are annealed directly adjacent to restriction sites of interest. Those paired at ligation junctions are ligated, enabling PCR amplification of the ligation products using the adapter sequences present within the primers. (E) Hi-C/ChIA-PET. All ligation junctions are labeled using a biotin tag. Complexes are purified using streptavidin, and adapters are ligated to the library to enable massive parallel sequencing.
suitable to obtain a detailed, yet comprehensive structure within a given gene locus.

More recently, methodologies have been devised to capture all interactions throughout the genome simultaneously. The Genome Conformation Capture (GCC) technique used by Rodley et al. [36] to capture yeast chromosome interactions simply involved next-generation sequencing of the entire 3C library, without selection or enrichment of ligation products. This is feasible due to the small genome size of yeast; however, GCC is unlikely to be applied in organisms with larger genomes. Three groups have designed methods to enrich for the 3C ligation junctions using biotinylated nucleotide or oligonucleotide tag sequences in between interacting ligated fragments, which is used to purify ligation junctions (Figure 1E). The Hi-C technique developed by Lieberman-Aiden et al. [37] incorporated biotin-based ligation selection to provide a population-average, whole-genome conformation snapshot of human nuclei with one megabase resolution, and gives insight into the organization of active and inactive chromatin domains, chromosome folding and preferred spatial arrangements of individual chromosome territories. The method applied by Duan et al. [38] was similar, where a modified circular 3C technique was combined with biotin enrichment to generate a three-dimensional model of the less complex haploid yeast genome with kilobase resolution, thus revealing the folding pattern of chromosomes and complexity of interchromosomal interactions. The third technique, ChIA-PET (chromatin interaction analysis with paired-end tag sequencing) differs most significantly from Hi-C and modified circular 3C by the method of generating fragments (sonication as opposed to restriction enzyme digestion), and its inclusion of an immunoprecipitation step to selectively identify the genomic loci interactions that occur within the context of a specific protein [39], much like ChiP Loop, 6C and e4C. Fullwood and colleagues used ChIA-PET to map the genomic interactions that co-associate with estrogen receptor alpha binding. This is an exciting development that may be capable of identifying interactions that take place at various nuclear subcompartments, which can provide clarification of the significance of many poorly understood nuclear bodies [40].

Of course, the utility of 3C derivatives to study nuclear organization would be limited had it not been for the genome sequencing projects to unambiguously map the positions of interacting loci, and powerful multiplexing technologies to capture a multitude of genomic interactions concurrently. The initial derivatives of 3C employed custom-made microarrays to assess the frequency of interactions [26, 35]. Gradually, this is being superseded by high-throughput next-generation sequencing, which can provide a greater versatility between experiments, and an ever-increasing output capacity. Certainly, such capacity will be needed to attain an adequate depth of coverage, given the extreme complexity of interactions that are likely to be present in Hi-C and ChIA-PET interaction libraries.

The development of these latest truly genome-wide interaction capture assays can offer a snapshot of whole-genome behavior in its entirety, and will be crucial to our understanding of structure–function relationships of the genome. However, these methods are unlikely to supersede e4C or 5C, which will provide a much more specific and detailed analysis of interactions from the perspective of a limited number of representative loci. Indeed, 3C will continue to be applied to study specific interactions, as well as to confirm those detected via the derivative methods. Akin to the use of different objective lenses on a microscope, the combination of methods can potentially provide a holistic view of genome organization, ranging from a detailed, localized view to an all-encompassing, global representation.

RNA FISH TRAP, which was used in parallel with 3C to capture β-globin gene promoter locus control region interactions, has not had such an immense impact on the genome organization studies as 3C. However, there is a considerable potential for its application to study unique aspects of nuclear organization. RNA FISH TRAP was used to investigate how the Air non-coding transcript represses other genes that reside within the locus [41]. The Air transcript was shown to localize over the promoter regions of these genes, and nucleate the recruitment of transcriptional repressors. This methodology can certainly be applied to other regulatory non-coding RNA species, such as the X chromosome-coating Xist transcript, to elucidate their mode of recruitment and function. Incorporation of next-generation sequencing of RNA FISH TRAP libraries may also ascertain the localization pattern of such non-coding RNA, to determine if interactions are restricted to genes in cis, or if they can occur in trans, as has been observed...
with the HOTAIR transcript [42], perhaps through nucleated non-coding RNA subcompartments. This new breed of techniques will open the door to understanding how and why the genome is organized, and provide insight into the roles of many nuclear subcompartments. However, it is important to be mindful of the limitations that accompany it. Libraries for 3C and its derivatives are typically prepared from tens of millions of cells, from which a population-averaged genome configuration will be obtained. While this is useful to explore many generalities of genomic interactions, the specific contexts of interactions such as alternate configurations and mutual exclusivity of interactions will be lost. Single-cell analysis may not yet be feasible for 3C applications, but analysis by microscopy may fill the niche. Developments in the field of light microscopy continue to push against the barriers of resolution limits, and may be capable of distinguishing alternate genomic conformations [43]. In addition, higher resolution electron spectroscopic imaging can be used to show how the genome interfaces with various nuclear subcompartments [44, 45]. Yet to truly understand the functional significance of genome interactions at both a local and global scale, it will be important to integrate these ‘interactomic’ datasets with the ever-accumulating transcriptomic (coding and non-coding RNA) and epigenomic (DNA and histone modifications) datasets, along with genome-wide binding profiles of key regulator proteins. An improved understanding of how they interface and influence each other should make clearer the causal relationships which will undoubtedly be key during cell fate decisions, development and disease.

Key Points
- Spatial organization of the genome has a considerable impact on nuclear function.
- Development of novel techniques for detecting chromatin interactions is rapidly advancing an understanding of functional nuclear organization.
- Coupling these techniques with the power of next-generation sequencing allows for the collection and collation of large datasets with which complex nuclear arrangements can be inferred.

Acknowledgements
We wish to thank S. Schoenfelder for useful comments in the manuscript. Cameron Osborne is a Leukemia and Lymphoma Research Bennett Fellow. Philip Ewels is supported by a BBSRC studentship, and Alice Young is supported by a MRC studentship.

References