From the chromatin interaction network to the organization of the human genome into replication N/U-domains

Rasha E Boulos\textsuperscript{1,2}, Hanna Julienne\textsuperscript{1,2}, Antoine Baker\textsuperscript{1,2}, Chun-Long Chen\textsuperscript{3}, Nataliya Petryk\textsuperscript{4}, Malik Kahli\textsuperscript{1}, Yves d’Aubenton-Carafa\textsuperscript{1}, Arach Goldar\textsuperscript{5}, Pablo Jensen\textsuperscript{1,2,6}, Olivier Hyrien\textsuperscript{4}, Claude Thermes\textsuperscript{3}, Alain Arneodo\textsuperscript{1,2} and Benjamin Audit\textsuperscript{1,2}

\textsuperscript{1}Université de Lyon, F-69000 Lyon, France
\textsuperscript{2}Laboratoire de Physique, ENS de Lyon, CNRS UMR5672, Lyon, F-69007, France
\textsuperscript{3}Centre de Génétique Moléculaire, CNRS UPR3404, Associée à l’Université Paris-Sud 11, FRC 3115, Gif sur Yvette, F-91198, France
\textsuperscript{4}Ecole Normale Supérieure, Institut de Biologie de l’ENS (IBENS), Inserm U1024, CNRS UMR 8197, Paris, F-75005, France
\textsuperscript{5}Commissariat à l’Energie Atomique, Ibitec-S, Gif sur Yvette, F-91191, France
\textsuperscript{6}IXXI, Rhône Alpes Institute of Complex Systems, F-69007 Lyon, France

E-mail: alain.arneodo@ens-lyon.fr and benjamin.audit@ens-lyon.fr

Received 1 July 2014, revised 11 September 2014
Accepted for publication 9 October 2014
Published 20 November 2014

New Journal of Physics \textbf{16} (2014) 115014
doi:10.1088/1367-2630/16/11/115014

Abstract

The three-dimensional (3D) architecture of the mammalian nucleus is now being unraveled thanks to the recent development of chromatin conformation capture (3C) technologies. Here we report the results of a combined multiscale analysis of genome-wide mean replication timing and chromatin conformation data that reveal some intimate relationships between chromatin folding and human DNA replication. We previously described megabase replication N/U-domains as mammalian multiorigin replication units, and showed that their borders are ‘master’ replication initiation zones that likely initiate cascades of origin firing responsible for the stereotypic replication of these domains. Here, we demonstrate that replication N/U-domains correspond to the structural domains of self-interacting chromatin, and that their borders act as insulating regions both in high-throughput 3C (Hi-C) data and high-resolution 3C (4C) experiments.
Further analyses of Hi-C data using a graph-theoretical approach reveal that N/U-domain borders are long-distance, interconnected hubs of the chromatin interaction network. Overall, these results and the observation that a well-defined ordering of chromatin states exists from N/U-domain borders to centers suggest that ‘master’ replication initiation zones are at the heart of a high-order, epigenetically controlled 3D organization of the human genome.

Keywords: chromatin interaction network, replication domains, graph theory, chromatin states, genome organization

1. Introduction

The three-dimensional (3D) dynamical architecture of eukaryotic genomes plays an important role in the regulation of nuclear functions [1–10]. The recent development of chromosome conformation capture (3C) technology [11] and its high-throughput extensions, 4C (circular 3C) [12, 13], 5C (3C carbon copy) [14], Hi-C (high-throughput 3C) [15], chIA-PET (chromatin interaction analyses by paired end tag sequencing) [16], and TCC (tethered conformation capture) [17] have led to rapid advances in the study of chromatin structure in yeast [11, 18], Drosophila [19, 20], mouse [12, 13, 21–24], and human [14–17, 22, 25]. These techniques provide quantitative measurement of the interaction frequency between two selected loci (3C), a selected locus and the rest of the genome (4C), multiple selected loci (5C), genome-wide pairs of loci (Hi-C, TCC), and genome-wide pairs of loci bound by a particular protein (chIA-PET). Because of cell-to-cell variation in chromatin folding, these chromatin interaction data have been considered as a proxy of some average 3D spatial proximity over a conformational ensemble. Pioneering works using 3C, 4C, and 5C technologies [12–14, 26] have shown that long-range looping interactions over tens or hundreds of kilobases are involved in gene regulation, such as between promoters, enhancers, and insulators. Genome-wide studies using high-resolution Hi-C technology have revealed that Drosophila [19, 20] and mammalian [22, 23] genomes are organized in megabase-sized, self-interacting structural domains localized on the chromosomes. Long-range interactions can bring together highly transcribed genes, offering a mechanism to explain the formation of transcription factories [2, 3, 27, 28]. Hi-C data have also been compared to genome-wide mean replication timing (MRT) data of related cell lines [29, 30], and consistent with the original Hi-C analysis of the human genome [15], a dichotomic picture has been proposed where early and late replication loci occur in separated compartments of open and closed chromatin, respectively.

As recently reviewed in [31], we described the megabase-sized replication ‘U/N-domains’ as multiorigin units of DNA replication of the mammalian genome [32–38]. When using a wavelet-based multiscale analysis of MRT data [39, 40] for seven human cell types, about half of the genome could be segmented in megabase-sized U-shaped MRT domains [37, 41–43], as illustrated in figures 1(a) and (c). Significant overlap was observed between the MRT U-domains of different cell types and also between U-domains and the germline replication domains previously defined as exhibiting an N-shaped nucleotide compositional skew profile [32–36, 44]. From the demonstration that the average fork polarity is directly reflected by both the compositional skew and the derivative of the MRT profile [37, 38, 45, 46], we argued that the N-shape of the latter in MRT U-domains sustains the existence of megabase-sized gradients of
Figure 1. (a) Mean replication timing profile (black curve) [40] from early 0 to late 1 along a 12-Mb-long fragment of human chromosome 1 in the GM06990 cell line. The horizontal colored bars correspond to four N/U-domains, identified in this fragment as MRT U-domains [37, 42] (red: 200-kb borders, blue: 400-kb center, green: interior). (b) Intrachromosome GM06990/NcoI Hi-C contact matrix [15] along the same fragment of human chromosome 1. Each pixel represents all interactions between pairs of 100-kb loci; intensity corresponding to the total number of reads is color-coded according to the color map. The dashed squares correspond to the four replication N/U-domains. (c) Average MRT profiles (±SEM) inside GM06990 MRT U-domains (blue); positions within each U-domain were rescaled between 0 and 1 so that the left (0) and right (1) borders are aligned. Numerical MRT (○, averaged over 6000 simulations) of the replication of a 1-Mb domain, obtained with the cascade model of origin firing [31, 48]. (d) Corresponding average $d\text{MRT}/dx$ profiles (±SEM) (blue) and mean replication fork polarity from simulations (○).
replication fork polarity in somatic and germline human cells (figure 1(d)) [37, 38, 42]. Recent sequencing of Okazaki fragments provides experimental confirmation of this linear, N-shaped change of fork polarity across MRT U-domains [47]. We collectively refer to these replication domains as N/U-domains [31]. We further proposed a model for the spatiotemporal program of replication along U-domains where, in addition to a small rate of random origin firing, replication preferentially initiates from the ‘master’ initiation zones at the N/U-domain borders, followed by a stochastic cascade of secondary initiations propagating to the domain center, possibly by fork-stimulated origin firing [31, 38]. Numerical simulations of a model including these features show that these ingredients sufficiently explain U-shaped MRT domains that present a linear gradient of fork polarity (figures 1(c) and (d)) [31, 48]. When investigating the large-scale organization of human genes inside these replication domains, a remarkable organization was revealed [34, 36]. Specifically, highly expressed genes in a given cell type are over-represented close to the corresponding N/U-domain borders [49]. Mapping experimental and numerical chromatin mark data in these domains shows that the ‘master’ replication initiation zones at the N/U-domain borders are specified by a region (~200 kb) of open and transcriptionally active chromatin [37, 50]. The additional comparative analysis of N/U-domains and Hi-C data has revealed that these functional domains are intimately linked to genome 3D architecture. In our previous work, we mainly reported results in the human erythroid cell line K562, using Hi-C data obtained via the HindIII restriction enzyme [37]. To assess the robustness of this link between functional domains and genome 3D architecture, we reproduce here these genome-wide analyses in human lymphoblastoid cell lines using a replication timing profile in GM06990 [40], Hi-C data obtained with the NcoI restriction enzyme (GM06990/NcoI) [15], and chromatin marks data in the related GM12878 cell line, gathered by the ENCODE project [51].

2. Replication N/U-domains and the 3D organization of a chromosome into structural domains

Replication occurs at discrete foci in the nucleus according to a specific spatiotemporal organization [52, 53]. Throughout the S phase, new replication foci de novo appear immediately next to the previously active foci [54–57], and a substantial spatial segregation between early- and late-replicated loci has been observed [58–61]. In addition, from the observation that thousands of replication forks are only found in hundreds of discrete BrdU-labeled foci, it has been proposed that DNA sequences replicating at the same time could gather [2, 5, 58, 62, 63], possibly through the anchoring of chromatin loops on the nuclear skeleton (reviewed in [41, 64, 65]). Moreover, other experiments showed that the spacing of anchoring sites on the nuclear skeleton increased with the replication fork velocity, and that the efficiency of an origin in a given S phase depended on its proximity to an anchorage point in the previous G1 phase [66]. Hence, there is a clear relationship between DNA replication and the pattern of the 3D chromatin interactions in the nucleus. The organization of the human spatiotemporal replication program into MRT U-domains [37] provided us with the opportunity to test this idea more systematically and to address the extent to which a tertiary chromatin-structure counterpart to these functional replication domains exists.

The original structural partitioning of the chromosomes derived from intrachromosomal Hi-C interactions (using a principal component analysis of the Hi-C data over each chromosome) and the normalization performed on the Hi-C data [15] gave as much weight
Figure 2. (a) Average number of GM06990/NcoI Hi-C interactions between 100-kb loci versus the distance separating them (logarithmic scales), as computed genome-wide (black) or restricted to pairs of loci belonging to same GM06990 N/U-domains, for four domain size categories: $L < 0.8\text{ Mb}$, $0.8\text{ Mb} < L < 1.2\text{ Mb}$, $1.2\text{ Mb} < L < 1.8\text{ Mb}$, and $1.8\text{ Mb} < L < 3\text{ Mb}$ (from blue to grey). (b) Ratio of the number of interactions between loci inside the same N/U-domain at equal distances from its center, and the number of interactions between loci on opposite sides and equal distances from an N/U-domain border, versus the distance between them for GM06990; the ratios were computed for the same four domain size categories as in (a). (c) Mean coverage by CTCF-enriched regions versus the distance to the closest N/U-domain border in the GM06990 (blue) and K562 (red) cell lines; for GM06990, CTCF data in GM12878 were used. Coverage values were normalized to the genome-wide mean values: 0.0046 in GM12878 and 0.0051 in K562. (d) 4C-like interaction profile of N/U-domain borders. We extracted Hi-C interaction profiles between N/U-domain borders and all their neighbors in the direction of the opposite N/U-domain border as a function of the distance to the reference border. To assure that the opposite N/U-domain border is well defined, we only kept 4C-like profiles where this border delimits two consecutive N/U-domains (N/U-domain 1 and 2). Interaction counts were normalized by the mean interaction count, averaged over all pairs of loci separated by the same genomic distance (in black in (a) for GM06990). All 4C-like interaction ratio profiles were finally averaged after rescaling the distances, so that all domain sizes were 1. Average over $n = 428$ 4C-like profiles in the K562 cell line (red) and $n = 465$ 4C-like profiles in the GM06990 cell line (blue).
to the very-distant interactions (over tens of megabases) as it gave to the more local contacts (below a few megabases), despite the fact that the latter are the most frequent contacts. Complementarily, we focused [37] on interactions between loci separated by short genomic distances ($\leq$10 Mb), over which the contact probabilities are the highest [15]. As illustrated in figure 1(b), it appears that replication N/U-domains correspond to matrix square-blocks of enriched interactions. This clearly suggests that the segmentation of the genome into replication N/U-domains coincides, to some extent, with the segmentation into structural domains identified by the analysis of more recent chromatin interaction data [22, 23]. Consistent with previous results, the matrix square-block structure underlines that the two early replicating zones that border a N/U-domain have high contact probability as the signature of 3D spatial proximity. However, the matrix square-block structure also signifies high contact probability of the two early replicating borders with the late replicating N/U-domain center, and sparse interactions between loci in separate N/U-domains. In other words, locally we do not observe a structural segregation between early- and late-replicating loci, and N/U-domain borders appear to correspond to structural insulating barriers. Further examination of the average behavior of intrachromosomal contact probability as a function of genomic distance for the complete genome corroborates these observations: the mean number of interactions between two 100-kb loci of the same N/U-domain decays when their genomic distance increases, as observed genome-wide (figure 2(a)). Importantly, the mean number of pairwise interactions is significantly higher inside the N/U-domains than genome-wide, and this seems to depend on the N/U-domain length. In particular, we found that the smaller the domain, the higher the mean number of interactions, which is probably a signature of a more open chromatin structure. When comparing the contact probability between loci inside an N/U-domain and loci lying in neighboring N/U-domains (figure 2(b)), we observed that their ratio increases with the distance separating the loci. Ratio values lower than one at short distances may be related to an excess of interactions in the open chromatin region at N/U-domain borders [37, 50]. Ratio values significantly greater than one for long distances corroborate the correspondence between structural barriers and N/U-domain borders. The latter observation is strengthened by the fact that N/U-domain borders are significantly enriched in the insulator-binding protein CTCF (figure 2(c)) [37], which is known to be involved in chromatin-loop formation, conditioning communication between transcriptional regulatory elements [67–70].

The Hi-C data do not always allow the direct visualization of the chromatin contacts for chosen loci, principally because of insufficient sequencing depth to cover all the combinatorial possibilities. This led us to use 4C methodology to explore the structural interaction partners of 10 viewpoints distributed along a large region (20 Mb) of human chromosome 5 [25]. The MRT profile of the region of interest presents large fluctuations with well-defined U-domains that are very well conserved across the seven cell lines we analyzed [37] (data not shown). In figure 3, we present 4C profiles for a viewpoint located in the middle of an N/U-domain and a viewpoint located on a well-defined timing peak at the border between two successive N/U-domains that are representative of the results for the 10 viewpoints analyzed [25]. Interaction frequency profiles are maximum at the viewpoint location, and quickly decrease to a few percents of the maximum value over a few megabases distance from the considered viewpoint. Importantly, this decrease does not follow a simple pattern. For the viewpoint located in the center of an N/U-domain, we observed that 4C interactions are mainly concentrated between the two timing peaks at the N/U-domain borders (figure 3, top). This very nicely confirms that contacts between late-replicating regions in N/U-domain central regions are somehow limited inside the
replication N/U-domains, and thus that there are self-interacting chromosomal domains bordered by timing peaks. Interestingly, beyond this local regime, late-replicated loci preferentially interact with other late-replicating regions (data not shown) [25], as suggested by the previous Hi-C data analysis. In contrast, for the viewpoint located at a timing peak, we observed peaks in the interaction frequency profile that colocalize with nearly every neighboring timing peak (figure 3, bottom), even with those located 10 Mb away. At larger distances, profiles of long-range interactions from timing peak viewpoints faithfully reproduce the replication timing profile (data not shown), which demonstrates the astonishing persistence, all over the chromosome, of a timing-linked long-range conformation of chromatin [25]. Each line of the Hi-C interaction matrix is equivalent to the 4C profile for the viewpoint situated on the diagonal, allowing us to computationally reproduce these analyses for all N/U-domains. We computed the average 4C-like profiles for viewpoints located at N/U-domain borders (see caption of figure 2(d)). We observed that both ends of the N/U-domains present preferential interactions (figure 2(d)); on average, there exists 1.5 (resp. 2) times more interactions between the two GM06990 (resp. K562) N/U-domain borders than expected, given their genomic distance (figure 2(d)). These results strongly suggest that N/U-domain borders loop out the intervening late-replicating region to contact each other. Note that all the results reported in figure 3 are robustly observed in cycling and G0-blocked cells, indicating that the observed organization likely reflects some permanent elements of the chromosome’s 3D structure.

Figure 3. 4C interaction profiles and replication timing in lymphoblastoid cells [25] along a 20-Mb region of human chromosome 5. Mean replication timing profile of the lymphoblastoid cell line GM06990 is shown in grey (early replication time at the top) [40]. Normalized 4C interaction frequencies profiles are shown for lymphoblastoids (red), resting peripheral blood mononuclear cells (PBMC) (green), and PHA-stimulated PBMC (blue). The arrowheads indicate the position of the viewpoints used for the 4C analyses: (top) V4 viewpoint is located at the mid-position of a MRT U-domain in GM06990 cell line, and (bottom) P1 viewpoint corresponds to a timing peak at the border between two consecutive N/U-domains in the GM06990 cell line. Adapted from Moindrot et al [25].
3. ‘Master’ replication initiation zones are long-range interconnected hubs in human chromatin interaction networks

These initial results prompted us to go further into the characterization of replication domains as fundamental units of the chromatin tertiary structure. We used graph theory to objectively quantify the importance of the ‘master’ replication initiation zones at N/U-domain borders in the genome-wide Hi-C chromatin network [71].

3.1. Graph theory tools to analyze interaction network complexity

Graphs [72] have become extremely useful as a representation of a wide variety of complex systems in social sciences, biology, computer sciences, and engineering [73–76]. An undirected graph consists of a set, $V$, of $n$ vertices, $v_i$, and a set, $E \subseteq V \times V$, of $m$ edges, $e_{ij}$, with associated weights, $w_{ij} > 0$. Note that unweighted graphs correspond to $w_{ij} = 1$. It is often useful to consider a matrix representation of a graph. The adjacency matrix, $A$, is an $n \times n$ square matrix whose entry, $a_{ij}$ ($i, j = 1, \ldots, n$), is equal to the weight, $w_{ij}$, and is zero otherwise. To identify and quantify the vertices that occupy critical positions in a network, centrality measures have been proposed, including the degree, betweenness, and eigenvector centralities [73, 76–79]. They can all be derived from the adjacency matrix. An obvious order of the vertices of a graph can be established by sorting them according to their degree, $d$. The corresponding degree centrality, $C_d$, is defined as [80]:

$$C_d(v_i) = \sum_{j=1}^{n} a_{ij}. \quad (1)$$

It is a local centrality measure, since it takes into account only the local structure (vertex neighborhood) of the graph. Betweenness centrality, $C_b$, measures the extent to which a vertex lies between other vertices on their geodesic paths [80]:

$$C_b(v_i) = \sum_{j,k \neq i}^{n} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}}, \quad (2)$$

where $\sigma_{jk}$ is the number of shortest paths* connecting $v_j$ and $v_k$, while $\sigma_{jk}(v_i)$ is the number of shortest paths connecting $v_j$ and $v_k$ and passing through $v_i$. A vertex with high betweenness centrality can potentially influence the spread of information through the graph by facilitating, hindering, or even altering the communication between other vertices. To distinguish vertices that are linked to well-connected vertices, and so influence many others in the graph either directly or indirectly through their connections, Bonacich [81] has proposed the following definition of the spectral centrality:

$$C_\lambda(v_i) = \sum_{j=1}^{n} a_{ij} C_\lambda(v_j). \quad (3)$$

This leads directly to the well-known problem of eigenvector computation, $\lambda C_\lambda = AC_\lambda$, and the eigenvector associated with the largest eigenvalue ($\lambda_{\text{max}}$) is the eigenvector centrality, $C_{\lambda_{\text{max}}}$. 

* Here, we define the length of a path $(e_{i_1j_1}, e_{i_2j_2}, \ldots, e_{i_kj_k})$ between vertices $v_i$ and $v_k$ as $L = \sum_{l=1}^{k} \frac{1}{w_{i_{l+1}j_l}}$. 

New J. Phys. 16 (2014) 115014 R E Boulos et al
Interaction maps are positively defined and symmetric, and so are prone to being represented and analyzed using graph theory [82]. A natural method involves the use of interaction matrices as the adjacency matrix, \( A \), of a graph, where the vertices, \( v_i \), are non overlapping DNA loci and the edges, \( e(v_i, v_j) \), link interacting loci. Two somehow complementary graph descriptions can be used: a weighted network where the weights assigned to each edge are the number of corresponding binary interactions, and a non weighted network where the entry, \( a_{ij} \), of \( A \) is 1 if \( v_i \) and \( v_j \) are interacting, and 0 if they are not interacting. Because the number of interactions decreases quickly when increasing the separation, \( s \), between the loci (like \( s^{-1} \)) [15, 82], the weighted network focuses on interactions between loci separated by short genomic distances (\( \leq 10 \) Mb) over which contact probabilities are the highest. Alternatively, the non weighted network equally takes into account short-range (\( \leq 10 \) Mb) and very-long-range interactions within a chromosome (\( \geq 40 \) Mb). In the latter case, we removed from the data all binary interactions that were present only once (\( t = 1 \)) or twice (\( t = 2 \), as some of these data may well be attributed to experimental noise (\( t = 0 \) corresponds to no thresholding).

3.2. Application of graph theory to Hi-C data

We applied the above methodology to the GM06990/NcoI Hi-C interaction matrix at the 100-kb resolution. For both the unweighted and weighted graph descriptions, replication domain borders are local maxima of the three centralities, confirming that they correspond to critical vertices in the Hi-C interaction network (figures 4(a)–(c)). The degree centrality, \( C_d \) (equation (1)), for the unweighted graph decreases almost linearly when moving inside the replication domains (figure 4(a)). As shown in figures 4(d) and (e), where the numbers of incident vertices to replication domain borders and centers are compared for different contact thresholds \( t = 0, 1, \) and 2, this decrease of \( C_d \) mainly results from the fact that borders have more long-range connections than centers (up to two-fold for \( n = 2 \)). When considering the weighted graph description, the relative decrease of \( C_d \) from border to center is much more pronounced than before (figure 4(a)), and this occurs regardless of the intragenomic length of the incident edges (\( > \) or \( \leq 4 \) Mb). This is the demonstration that replication domain borders have a much higher intrachromosomal contact frequency with other short-distance or long-distance loci than domain centers [37]. In figure 4(b), the betweenness centrality, \( C_b \) (equation (2)), is also shown to decrease faster for the weighted representation than the unweighted one. This quantifies the role of these replication domain borders as ‘hubs’ in the chromatin interaction network. These hubs are not only important actors in mediating long-range interactions (\( > 4 \) Mb), they also play a fundamental role at short distances (\( \leq 4 \) Mb), where the very sharp decrease of \( C_b \) observed in figure 4(b) enlightens the insulator properties of the replication domain borders that prevent cross-talk between neighboring domains, likely establishing self-interacting, independent expression domains [37]. In figure 4(c), the eigenvector centrality, \( C_\lambda \) (equation (3)), also decreases from replication domain borders to centers. The comparison of the relative decay obtained with the weighted and unweighted graph representations further strengthens the fundamental organizing role of the replication domain borders, which are hubs that predominantly interact at short distances (\( \leq 4 \) Mb) with nearby hubs as a possible signature of 3D spatial proximity. The results reported in figure 4(c) also show that these hubs have preferential long-range interactions with other very distal hubs. To visualize these interaction graph properties, we present in figure 5 the results of a simple, two-dimensional (2D), particle-
like model where genomic loci both attract each other like springs based on their observed interactions, and repel each other like magnets to avoid over-aggregation [71]. Genomic loci are arranged in a worm-like fashion following their ordering along the chromosome due to the quick decay in contact frequency as a function of genomic distance [15]. Replication domain borders are found closer to each other in the graph layout obtained from this particle model, unlike centers that mainly interact at short distances and are found looping out at the periphery.

Figure 4. (a) Degree, (b) betweenness, (c) eigenvector centralities averaged over the 1020 replication N/U-domains delineated in the GM06990 cell line versus the distance to the closest border: unweighted graph (◯), weighted graph for all edges (△), or limited to edges connecting loci separated by more than 4 Mb (—), or less than 4 Mb (- - -). For the sake of comparison, \( \langle C_d \rangle \) (equation (1)), \( \langle C_i \rangle \) (equation (2)) and \( \langle C_{\lambda} \rangle \) (equation (3)) are represented relative to their maximal value. (d) Probability of interaction between 100-kb loci at replication domain borders (resp. centers) with other 100-kb loci on the same chromosome versus their intragenomic distance; interactions with the two 100-kb loci at the each N/U-domain border (red) and the four 100-kb loci at the domain centers (blue) are considered. (e) Ratio between the probabilities of interaction of replication domain borders and centers (shown in (d)) versus the intragenomic interaction distances for different contact thresholds \( t = 0 \) (◯), 1(□), and 2 (△). To better discriminate centers from borders, only replication domains longer than 2 Mb were taken into consideration in (d) and (e).
of this stationary, worm-like final configuration, which illustrates the central role of the interaction network hubs in the specification of the 3D architecture of the human chromosome. Recent analyses of the network of interaction between human organ systems showed that there are different network topologies corresponding to different physiological states, and that pairs of systems can communicate and synchronize their function through simultaneously coexisting forms of coupling [83–85]. It will be interesting to see if the concepts introduced in network physiology apply when considering the chromatin-mediated coregulation of nuclear functions.

4. Organization of chromatin states along replication N/U-domains

There are tens of epigenetic marks that can be either present or absent on a chromatin locus, suggesting that the number of distinct chromatin states is huge. Several independent analyses [19, 87–89] have shown that most epigenetic mark combinations in human, Drosophila melanogaster, and Arabidopsis thaliana are not (or are only minimally) observed, so that the number of distinct epigenetic states is in fact very limited (≤15). In order to go beyond our previous analysis of chromatin mark distribution in replication domains that consisted of computing the average profile of their presence/absence as a function of the distance to the domain borders [37, 50], we recently followed these pioneering approaches to more precisely characterize the link between the mean replication timing and the epigenetic status. Using principal component analysis [90] and classical clustering [91] on a set of 13 epigenetic marks at a spatial resolution of 100 kb (corresponding to MRT profile resolution), we reduced the apparent complexity of the data set to four major groups of chromatin marks with shared features [92]. The small number of states found, compared to [88] who described 15 states in human, is likely due to the much larger resolution used in this study. These states have specific replication timing, an early transcriptionally active euchromatin state (C1), a mid-early
Figure 6. Distribution of the four chromatin states determined in GM12878 inside GM06990 replication-timing N/U-domains. (a) The N/U-domains were centered and ordered vertically from the smallest (top) to the largest (bottom). All transcriptionally active chromatin state C1 100-kb windows were represented by a horizontal segment of the corresponding length. (b) Same as (a), for the Pc repressed chromatin state C2. (c) Same as (a), for the silent, unmarked chromatin state C3. (d) Same as (a), for the HP1 heterochromatin state C4. (e) Mean coverage of chromatin state, with respect to the distance to the closest N/U-domain border for N/U-domains smaller than 0.8 Mb. Error bars represent the standard deviation of the mean. (f) Same as (e), for N/U-domains between 0.8 Mb and 1.2 Mb in size. (g) Same as (e), for N/U-domains between 1.2 Mb and 1.8 Mb in size. (h) Same as (e), for N/U-domains between 1.8 Mb and 3.0 Mb in size. Colors correspond to the four chromatin states: transcriptionally active chromatin (pink), chromatin repressed by polycomb (orange), silent unmarked chromatin (green), and HP1 heterochromatin (blue).
repressive type of chromatin (C2) associated with polycomb complexes, and two late replicating states: a silent state (C3) not enriched in any available marks, and a gene-poor HP1-associated heterochromatin state (C4). When mapping these chromatin states inside the megabase-sized N/U-domains, it appears that the associated replication-fork polarity gradient [37, 38] corresponds to a directional path across three chromatin states, from C1 at the N/U-domain borders followed by C2 and C3 at the centers (figure 6). For the lymphoblastoid data analyzed here, state C4 is rather evenly distributed along the N/U-domains, which contrasts with the situation in the K562 cell line, where C4 is mainly localized in the central regions of the largest domains (≥2 Mb) [92]. Thus, in K562, the gradient of chromatin states involves the four states from C1 to C4. Interestingly, analysis of the half of the human genome not covered by N/U-domains is consistent with early and late replication loci occurring in separate compartments; the former correspond to gene-rich, high-GC domains of intermingled chromatin states C1 and C2, whereas the latter correspond to gene-poor, low-GC domains of alternating chromatin states C3 and C4, or long C4 domains. The relationship between the distribution of chromatin states, the spatiotemporal replication program, and the 3D genome architecture could be the key to understanding the chromatin complex organization [93] and its role in regulating nuclear functions [94, 95].

5. Conclusion

To summarize, the organization of the human genome into replication domains covering about half of the genome provides us with an original point of view to unify the interplay between replication, transcription, epigenetic modifications, and nuclear organization. ‘Master’ replication initiation zones at N/U-domain borders are major actors in the spatiotemporal replication program [31, 37, 38, 50], acting as insulating regions between chromatin structural domains of independent expression and duplication [25, 37, 71]. These regions of open and transcriptionally active chromatin also play a central role in the genome’s 3D architecture by mediating long-range interactions among distant DNA elements both within and in between chromosomes. This property is consistent with the observation that N/U-domain borders are significantly enriched in the insulator-binding protein CTCF (figure 2(c)), which may contribute to the formation of chromosomal hubs of interactions across chromosomes [82]. Analysis of the spatial proximity of evolutionary breakpoints between human and mouse suggests that some aspects of genome 3D architecture are conserved across very large evolutionary distances [96, 97]. This suggests that replication/structural domains could be major determinants of genome evolution, which is consistent with the observation that some specific replication timing patterns are conserved between human and mouse syntenic regions of related cell types despite the length of evolutionary divergence [30], and with our previous finding that replication N/U-domain borders are significantly enriched in mammalian evolutionary breakpoints, which suggests that evolution typically shuffled these structural and functional domains rather than breaking and fusing them [96]. Replication/structural domains are also likely central to genome regulation on shorter time scales. Comparative analyses of replication timing profiles during cell differentiation in mammals underline that there are important dynamical changes of replication timing profiles during development, leading to cell-type specific patterns of replication [30, 95, 98, 99]. Meta-analyses [100, 101] of replication timing profiles [40], Hi-C data [15], and somatic copy-number alterations (SCNAs) observed in cancer samples from diverse cancer
types [102] showed that SCNAs tend to fuse genomic regions that, prior to the rearrangement, spatially co-localized within the nucleus and have similar replication timing. Hence, replication N/U-domains likely correspond to fundamental structural and functional units underlying the plasticity of replication and structural domain organization, gene expression, and chromatin states both across different cell lines and across different organisms. They provide a framework for further studies in different cell types, in both health and disease.

Acknowledgments

This work was supported by the Agence Nationale de la Recherche (ANR), project REFOPOL, (ANR 10 BLAN 1615).

References

[1] Belmont A S, Dietzel S, Nye A C, Strukov Y G and Tumbar T 1999 Large-scale chromatin structure and function *Curr. Opin. Cell Biol.* 11 307–11
[2] Cook P R 1999 The organization of replication and transcription *Science* 284 1790–5
[3] Cook P R 2001 *Principles of Nuclear Structure and Functions* (New York: Wiley)
[4] Cremer T and Cremer C 2001 Chromosome territories, nuclear architecture and gene regulation in mammalian cells *Nat. Rev. Genet.* 2 292–301
[5] Berezney R 2002 Regulating the mammalian genome: the role of nuclear architecture *Adv. Enzyme Regul.* 42 39–52
[6] Gilbert N, Gilchrist S and Bickmore W A 2005 Chromatin organization in the mammalian nucleus *Int. Rev. Cytol.* 242 283–336
[7] Misteli T 2007 Beyond the sequence: cellular organization of genome function *Cell* 128 787–800
[8] Sexton T, Schober H, Fraser P and Gasser S M 2007 Gene regulation through nuclear organization *Nat. Struct. Mol. Biol.* 14 1049–55
[9] Branco M R and Pombo A 2007 Chromosome organization: new facts, new models *Trends Cell. Biol.* 17 127–34
[10] Fraser P and Bickmore W 2007 Nuclear organization of the genome and the potential for gene regulation *Nature* 447 413–7
[11] Dekker J, Rippe K, Dekker M and Kleckner N 2002 Capturing chromosome conformation *Science* 295 1306–11
[12] 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 1348–54
[13] Zhao Z et al 2006 Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions *Nat. Genet.* 38 1341–7
[14] Dostie J et al 2006 Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements *Genome Res.* 16 1299–309
[15] Lieberman-Aiden E et al 2009 Comprehensive mapping of long-range interactions reveals folding principles of the human genome *Science* 326 289–93
[16] Fullwood M J et al 2009 An oestrogen-receptor-alpha-bound human chromatin interactome *Nature* 462 58–64
[17] Kalhor R, Tjong H, Jayathilaka N, Alber F and Chen L 2012 Genome architectures revealed by tethered chromosome conformation capture and population-based modeling *Nat. Biotechnol.* 30 90–98
[18] Duan Z et al 2010 A three-dimensional model of the yeast genome *Nature* 465 363–7
[19] Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A and Cavalli G 2012 Three-dimensional folding and functional organization principles of the drosophila genome Cell 148 458–72

[20] Hou C, Li L, Qin Z S and Corces V G 2012 Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains Mol. Cell 48 471–84

[21] Zhang Y, McCord R P, Ho Y-J, Lajoie B R, Hildebrand D G, Simon A C, Becker M S, Alt F W and Dekker J 2012 Spatial organization of the mouse genome and its role in recurrent chromosomal translocations Cell 148 908–21

[22] Dixon J R, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu J S and Ren B 2012 Topological domains in mammalian genomes identified by analysis of chromatin interactions Nature 485 376–80

[23] Nora E P et al 2012 Spatial partitioning of the regulatory landscape of the X-inactivation centre Nature 485 381–5

[24] Takebayashi S I, Dileep V, Ryba T, Dennis J H and Gilbert D M 2012 Chromatin-interaction compartment switch at developmentally regulated chromosomal domains reveals an unusual principle of chromatin folding Proc. Natl Acad. Sci. USA 109 12574–9

[25] Moindrot B, Audit B, Klous P, Baker A, Thermes C, de Laat W, Bouvet P, Mongelard F and Arneodo A 2012 3D chromatin conformation correlates with replication timing and is conserved in resting cells Nucleic Acids Res. 40 9470–81

[26] Splinter E and de Laat W 2011 The complex transcription regulatory landscape of our genome: control in three dimensions EMBO J. 30 4345–55

[27] Carter D R F, Eskiw C and Cook P R 2008 Transcription factories Biochem. Soc. Trans. 36 585–9

[28] Schoenfelder S et al 2010 Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells Nat. Genet. 42 53–61

[29] Yaffe E, Farkash-Amar S, Polten A, Yakhini Z, Tanay A and Simon I 2010 Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture PLoS Genet. 6 e1001011

[30] Ryba T et al 2010 Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types Genome Res. 20 761–70

[31] Hyrien O et al 2013 From simple bacterial and archaeal replicons to replication N/U-domains J. Mol. Biol. 425 4673–89

[32] Brodie of Brodie E-B, Nicolay S, Touchon M, Audit B, d’Aubenton-Cara Y, Thermes C and Arneodo A 2005 From DNA sequence analysis to modeling replication in the human genome Phys. Rev. Lett. 94 248103

[33] Touchon M, Nicolay S, Audit B, Brodie of Brodie E-B, d’Aubenton-Cara Y, Arneodo A and Thermes C 2005 Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins Proc. Natl Acad. Sci. USA 102 9836–41

[34] Huvet M, Nicolay S, Touchon M, Audit B, d’Aubenton-Cara Y, Arneodo A and Thermes C 2007 Human gene organization driven by the coordination of replication and transcription Genome Res. 17 1278–85

[35] Audit B, Nicolay S, Huvet M, Touchon M, d’Aubenton-Cara Y, Thermes C and Arneodo A 2007 DNA replication timing data corroborate in silico human replication origin predictions Phys. Rev. Lett. 99 248102

[36] Baker A, Nicolay S, Zaghoul L, d’Aubenton-Cara Y, Thermes C, Audit B and Arneodo A 2010 Wavelet-based method to disentangle transcription- and replication-associated strand asymmetries in mammalian genomes Appl. Comput. Harmon. Anal. 28 150–70

[37] Baker A et al 2012 Replication fork polarity gradients revealed by megabase-sized U-shaped replication timing domains in human cell lines PLoS Comput. Biol. 8 e1002443

[38] Guilbaud G et al 2011 Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome PLoS Comput. Biol. 7 e1002322

[39] Chen C-L et al 2010 Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes Genome Res. 20 447–57
Hansen R S, Thomas S, Sandstrom R, Canfield T K, Thurman R E, Weaver M, Dorschner M O, Gartner S M and Stamatoyannopoulos J A 2010 Sequencing newly replicated DNA reveals widespread plasticity in human replication timing Proc. Natl Acad. Sci. USA 107 139–44

Arneodo A, Vaillant C, Audit B, Argoul F, d’Aubenton-Carafa Y and Thermes C 2011 Multi-scale coding of genomic information: From DNA sequence to genome structure and function Phys. Rep. 498 45–188

Audit B, Baker A, Chen C-L, Rappailles A, Guilbaud G, Julienne H, Goldar A, d’Aubenton Carafa Y and Hyrien O 2013 Multiscale analysis of genome-wide replication timing profiles using a wavelet-based signal-processing algorithm Nat. Protoc. 8 98–110

Audit B, Zaghloul L, Baker A, Arneodo A, Chen C-L, d’Aubenton Carafa Y and Thermes C 2012 Megabase replication domains along the human genome: relation to chromatin structure and genome organisation Subcell. Biochem 61 57–80

Chen C-L et al 2011 Replication-associated mutational asymmetry in the human genome Mol. Biol. Evol. 28 2327–37

Baker A, Julienne H, Chen C L, Audit B, d’Aubenton-Carafa Y, Thermes C and Arneodo A 2012 Linking the DNA strand asymmetry to the spatiotemporal replication program. I. About the role of the replication fork polarity in genome evolution Eur. Phys. J. E 35 92

Baker A, Chen C L, Julienne H, Audit B, d’Aubenton-Carafa Y, Thermes C and Arneodo A 2012 Linking the DNA strand asymmetry to the spatiotemporal replication program: II. Accounting for neighbor-dependent substitution rates Eur. Phys. J. E 35 123

Petryk N et al 2011 Replication fork directionality of the human genome (in preparation)

Goldar A 2014 private communication

Zaghloul L, Baker A, Audit B and Arneodo A 2012 Gene organization inside replication domains in mammalian genomes C. R. Mécanique 340 745–57

Audit B, Zaghloul L, Vaillant C, Chevereau G, d’Aubenton-Carafa Y, Thermes C and Arneodo A 2009 Open chromatin encoded in DNA sequence is the signature of ‘master’ replication origins in human cells Nucleic Acids Res. 37 6064–75

The ENCODE Project Consortium 2007 Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project Nature 447 799–816

Nakayasu H and Berezney R 1989 Mapping replicational sites in the eucaryotic cell nucleus J. Cell Biol. 108 1–11

O’Keefe R T, Henderson S C and Spector D L 1992 Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences J. Cell Biol. 116 1095–110

Leonhardt H, Rahn H P, Weinzierl P, Sporbert A, Cremer T, Zink D and Cardoso M C 2000 Dynamics of DNA replication factories in living cells J. Cell Biol. 149 271–80

Sporbert A, Gahl A, Ankerhold R, Leonhardt H and Cardoso M C 2002 DNA polymerase clamp shows little turnover at established replication sites but sequential de novo assembly at adjacent origin clusters Mol. Cell 10 1355–65

Sadoni N, Cardoso M C, Stelzer E H K, Leonhardt H and Zink D 2004 Stable chromosomal units determine the spatial and temporal organization of DNA replication J. Cell Sci. 117 5353–65

Maya-Mendoza A, Olivares-Chauvet P, Shaw A and Jackson D A 2010 S phase progression in human cells is dictated by the genetic continuity of DNA foci PLoS Genet. 6 e1000900

Ma H, Samarabandu J, Devdhari R S, Acharya R, Cheng P C, Meng C and Berezney R 1998 Spatial and temporal dynamics of DNA replication sites in mammalian cells J. Cell Biol. 143 1415–25

Zink D, Cremer T, Saffrich R, Fischer R, Trendelenburg M F, Ansonge W and Stelzer E H 1998 Structure and dynamics of human interphase chromosome territories in vivo Hum. Genet. 102 241–51

Zink D, Bornfleth H, Visser A, Cremer C and Cremer T 1999 Organization of early and late replicating DNA in human chromosome territories Exp. Cell Res. 247 176–88
[61] Grasser F, Neusser M, Fiegler H, Thomeyer T, Cremer M, Carter N P, Cremer T and Müller S 2008 Replication-timing-correlated spatial chromatin arrangements in cancer and in primate interphase nuclei J. Cell Sci. 121 1876–86

[62] Jackson D A and Pombo A 1998 Replicon clusters are stable units of chromosome structure: evidence that nuclear organization contributes to the efficient activation and propagation of S phase in human cells J. Cell Biol. 140 1285–95

[63] Koberna K et al 2005 Electron microscopy of DNA replication in 3-D: evidence for similar-sized replication foci throughout S-phase J. Cell. Biochem. 94 126–38

[64] Anachkova B, Djeliova V and Russev G 2005 Nuclear matrix support of DNA replication J. Cell. Biochem. 96 951–61

[65] Ottaviani D, Lever E, Takousis P and Sheer D 2008 Anchoring the genome Genome Biol. 9 201

[66] Courbet S, Gay S, Arnoult N, Wronka G, Anglana M, Brison O and Debatisse M 2008 Replication fork movement sets chromatin loop size and origin choice in mammalian cells Nature 455 557–60

[67] Phillips J E and Corces V G 2009 The structure and function of complex networks SIAM Rev. 45 167–256

[68] Houb C, Dale R and Dean A 2010 CTCF: master weaver of the genome Nat. Rev. Genet. 5 101–13

[69] Scott J 2000 Social Network Analysis: A Handbook 2nd ed (London: Sage Publications)

[70] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D 2006 Complex networks: structure and dynamics Phys. Rep. 424 175

[71] Albert R and Barabasi A-L 2002 Statistical mechanics of complex networks Rev. Mod. Phys. 74 47

[72] Caldarelli G and Vespignani A (ed) 2007 Large Scale Structure and Dynamics of Complex Networks (Singapore: World Scientific)

[73] Scott J 2000 Social Network Analysis: A Handbook 2nd ed (London: Sage Publications)

[74] Albert R and Barabasi A-L 2002 Statistical mechanics of complex networks Rev. Mod. Phys. 74 47

[75] Caldarelli G and Vespignani A (ed) 2007 Large Scale Structure and Dynamics of Complex Networks (Singapore: World Scientific)

[76] Freeman L S 1979 Centrality in social networks: conceptual clarification Social Networks 1 215

[77] Bonacich P 1972 Factoring and weighting approaches to status scores and clique identification J. Math. Sociol. 2 113

[78] Botta M, Haider S, Leung I X Y, Lio P and Mozziiconacci J 2010 Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide Mol. Syst. Biol. 6 426

[79] Bashan A, Bartsch R P, Kantelhardt J W, Havlin S and Ivanov P C 2012 Network physiology reveals relations between network topology and physiological function Nat. Commun. 3 702

[80] Ivanov P C and Bartsch R P 2014 Network physiology: mapping interactions between networks of physiologic networks Networks of Networks: The Last Frontier of Complexity ed G D’Agostino and A Scala (Switzerland: Springer International Publishing) pp 203–22

[81] Ivanov P C and Bartsch R P 2014 Network physiology: mapping interactions between networks of physiologic networks Networks of Networks: The Last Frontier of Complexity ed G D’Agostino and A Scala (Switzerland: Springer International Publishing) pp 203–22

[82] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D 2006 Complex networks: structure and dynamics Phys. Rep. 424 175

[83] Albert R and Barabasi A-L 2002 Statistical mechanics of complex networks Rev. Mod. Phys. 74 47

[84] Caldarelli G and Vespignani A (ed) 2007 Large Scale Structure and Dynamics of Complex Networks (Singapore: World Scientific)

[85] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D 2006 Complex networks: structure and dynamics Phys. Rep. 424 175

[86] Bastian M, Heymann S and Jacomy M 2009 Gephi: An open source software for exploring and manipulating networks Third International AAAI Conf. on Weblogs and Social Media (www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154.h)
Filion G J et al 2010 Systematic protein location mapping reveals five principal chromatin types in drosophila cells Cell 143 212–24

Ernst J et al 2011 Mapping and analysis of chromatin state dynamics in nine human cell types Nature 473 43–49

Roudier F et al 2011 Integrative epigenomic mapping defines four main chromatin states in Arabidopsis EMBO J. 30 1928–38

Izenman A J 2008 Modern Multivariate Statistical Techniques: Regression Classification, and Manifold Learning (New York: Springer)

Kaufman L and Rousseeuw P J 1984 Finding Groups in Data: An Introduction to Cluster Analysis (New York: John Wiley & Sons)

Julienne H, Zoufir A, Audit B and Arneodo, A 2013 Human genome replication proceeds through four chromatin states PLoS Comput. Biol. 9 e1003233

Barbieri M, Chotalia M, Fraser J, Lavitas L-M, Dostie J, Pombo A and Nicodemi M 2012 Complexity of chromatin folding is captured by the strings and binders switch model Proc. Natl Acad. Sci. USA 109 16173–8

Julienne H, Zoufir A, Audit B and Arneodo A 2013 Epigenetic regulation of the human genome: coherence between promoter activity and large-scale chromatin environment Front. Life Sci. 7 44–62

Julienne H, Audit B and Arneodo A 2014 Embryonic stem cell specific ‘master’ replication origins at the heart of the loss of pluripotency PLoS Comput. Biol. doi:10.1371/journal.pcbi.1003969

Lemaitre C, Zaghloul L, Sagot M-F, Gautier C, Arneodo A, Tannier E and Audit B 2009 Analysis of fine-scale mammalian evolutionary breakpoints provides new insight into their relation to genome organisation BMC Genomics 10 335

Véron A S, Lemaitre C, Gautier C, Lacroix V and Sagot M-F 2011 Close 3D proximity of evolutionary breakpoints argues for the notion of spatial synteny BMC Genomics 12 303

Hiratani I et al 2008 Global reorganization of replication domains during embryonic stem cell differentiation PLoS Biol. 6 e245

Ryba T, Hiratani I, Sasaki T, Battaglia D, Kulik M, Zhang J, Dalton S and Gilbert D M 2011 Replication timing: a fingerprint for cell identity and pluripotency PLoS Comput Biol 7 e1002225

De S and Michor F 2011 DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes Nat. Biotechnol. 29 1103–8

Fudenberg G, Getz G, Meyerson M and Mirny L A 2011 High order chromatin architecture shapes the landscape of chromosomal alterations in cancer Nat. Biotechnol. 29 1109–13

Beroukhim R et al 2010 The landscape of somatic copy-number alteration across human cancers Nature 463 899–905