

Review

The “Genomic Code”: DNA Pervasively Moulds Chromatin Structures Leaving no Room for “Junk”

Giorgio Bernardi ^{1,2} 

¹ Science Department, Roma Tre University, Viale Marconi 446, 00146 Rome, Italy; gbernardi@uniroma3.it; Tel.: +39-33-540-5892

² Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy

Abstract: The chromatin of the human genome was analyzed at three DNA size levels. At the first, compartment level, two “gene spaces” were found many years ago: A GC-rich, gene-rich “genome core” and a GC-poor, gene-poor “genome desert”, the former corresponding to open chromatin centrally located in the interphase nucleus, the latter to closed chromatin located peripherally. This bimodality was later confirmed and extended by the discoveries (1) of LADs, the Lamina-Associated Domains, and InterLADs; (2) of two “spatial compartments”, A and B, identified on the basis of chromatin interactions; and (3) of “forests and prairies” characterized by high and low CpG islands densities. Chromatin compartments were shown to be associated with the compositionally different, flat and single- or multi-peak DNA structures of the two, GC-poor and GC-rich, “super-families” of isochores. At the second, sub-compartment, level, chromatin corresponds to flat isochores and to isochore loops (due to compositional DNA gradients) that are susceptible to extrusion. Finally, at the short-sequence level, two sets of sequences, GC-poor and GC-rich, define two different nucleosome spacings, a short one and a long one. In conclusion, chromatin structures are moulded according to a “genomic code” by DNA sequences that pervade the genome and leave no room for “junk”.

Keywords: Genomic code; chromatin structure; junk DNA



Citation: Bernardi, G. The “Genomic Code”: DNA Pervasively Moulds Chromatin Structures Leaving no Room for “Junk”. *Life* **2021**, *11*, 342. <https://doi.org/10.3390/life11040342>

Academic Editor: Lluís Ribas de Pouplana, Michael Russell, Dirk Schulze-Makuch, Michael W. Friedrich, Kousuke Hanada, Stefanos Roumeliotis and Hiroki Nishikawa

Received: 1 March 2021

Accepted: 7 April 2021

Published: 13 April 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The term “genome” was coined one hundred years ago to define “the haploid chromosome ensemble together with the associated protoplasm”, by Hans Winkler [1], a Professor of Botany at the University of Hamburg (and a frequent visitor at the Stazione Zoologica of Naples; Christiane Groeben, personal comm.). This definition became popular only after the discoveries of three pillars of the human genome, the double helix [2], the regulatory sequences [3], and the genomic code [4]. While the double helix concerned the totality of the genome, coding and regulatory sequences occupied very small percentages of it, leaving vast spaces in the genome that were reasonably proposed to correspond to “junk DNA” [5]. Ohno identified this fraction of the genome to accommodate for large numbers of deleterious mutations that would otherwise result in heavy mutation load, if they were in functional regions of the genome [5]. In this study, however, we will argue that this portion of the genome is functional, and thus is not to be considered as junk, yet not coding for proteins, and as such can withstand mutations without altering their function.

This article (which was written during the COVID pandemic, when the University was closed) presents a new general picture of the organization of the human genome based on DNA sequences. It is focused on three major topics: (1) The old discovery of the bimodal compositional compartmentalization of the human genome and a comparison with later results from other approaches; (2) Recent investigations that link GC-poor and GC-rich DNA sequences characterized by compositionally flat and single- or multi-peak structures, respectively, to the long-range chromatin compartments, to the short-range chromatin sub-compartments corresponding to flat isochores and isochore loops, and to the short

sequences responsible for nucleosome binding; and (3) The “genomic code”, the fourth pillar of the genome, which underlies the moulding of chromatin by DNA sequences; while the “genomic code” was discovered four years ago and discussed [6–9], a more complete picture will be presented here.

2. Isochores, Chromosomal Bands, Replication Timings, Evolutionary Transitions

The structural approach to the study of the genome, a top-down approach initiated sixty years ago [10], led to an important discovery: the demonstration of “major DNA components” in the mammalian genome [11], which, incidentally, occurred very close in time with the “junk DNA” proposal [5]. Indeed, (1) it disposed of the then generally accepted view that the DNA of higher organisms showed a continuous variation in GC level [12]; and (2) proposed that the genome of vertebrates is “a mosaic of isochores” [11,13–15], as we called the “major DNA components”, namely the “fairly homogeneous sequences” that belong, in the case of the human genome, in five families, L1, L2, H1, H2, H3, characterized by increasing GC levels (Figure 1A). The human genome was later estimated to comprise ~3200 isochores having an average size of 0.9 Mb [16]. More recently, (1) the human genome melting map was shown to co-vary very strongly with GC levels [17]; and (2) an approach, “isosegmenter”, was developed which allowed segmenting vertebrate genomes into isochores in a fast and completely automatic manner [18].

Three lines of evidence connect isochores with important genome structures and with genome evolution.

Our early proposal that the DNAs of Giemsa and Reverse bands correspond to GC-poor and GC-rich isochores, respectively [15], was followed by many investigations that defined the correlations between isochores and chromosomal bands from early prophase to metaphase [19–22].

Mammalian chromosomes were known to be replicated from many origins, adjacent starting points initiating replication at the same time, early and late in Reverse and Giemsa bands, respectively. More recent investigations confirmed the expected correlation between replication timing and isochore composition of the bands [23–25]. In fact, replicons located in a given isochore always show either all early or all late replication timings; moreover, early-replication isochores are short and GC-rich, late-replicating isochores long and GC-poor [21].

A third important point concerned the GC enrichment of isochores at the emergence of warm-blooded vertebrates. This stabilized thermodynamically at the higher body temperature of the latter a large part of the genome from cold-blooded vertebrates (a smaller part did not change its base composition) and was responsible for the compositional bimodal structure of the genome of warm-blooded vertebrates [26–30]. Moreover, it supported the idea that GC-rich isochores arose from adaptive evolution [19,29]. Interestingly, some fishes living at high temperatures showed GC-rich isochores that were absent in evolutionarily close species living at lower temperatures [26]. This GC-specific enrichment was also found in coding sequences of some turtles and crocodiles [31–34].

Finally, it should be mentioned that the initial sequencing and analysis of the human genome published by the International Human Genome Sequencing Consortium, IHGSC (2001) [35] comprised a number of misunderstandings about isochores that were discussed and clarified [36]. These problems seem now to be solved (see Section 11.4).

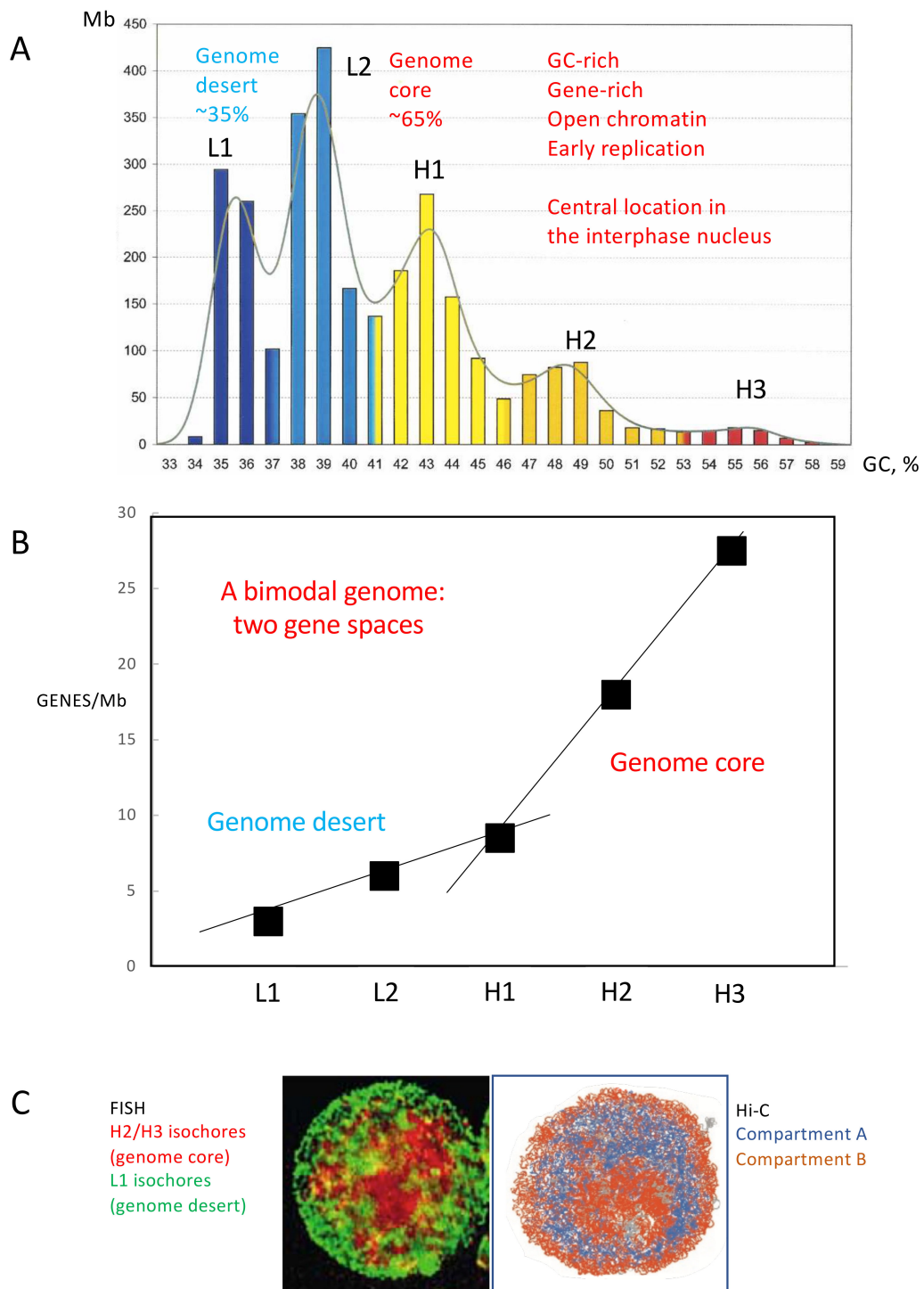


Figure 1. (A) Isochore families. The histogram displays the isochores from the human genome as pooled in bins of 1% GC. The Gaussian profile shows the distribution of isochore families, which are represented in different colours. Gene densities define a “genome desert”, comprising isochore families L1 and L2⁻, and a “genome core”, comprising isochore families L2⁺, H1, H2, H3; L2⁻ and L2⁺ sub-families are separated by a vertical broken red line. (Modified from [37]). (B) Gene distribution in the isochore families of the bovine genome. (Drawn from data of [37]); Two different lines of gene concentration characterized by different slopes were drawn through L1-H1 and H1-H3 points. (C) A FISH pattern of H2/H3 and L1 isochores (a clearer version of Figure 2A3 of [38]) is compared with a Hi-C pattern (modified from [39]), in which case compartment A is central, compartment B is peripheral and nucleolar.

3. Sequence Distribution in the Human Genome and an Early View of the Genomic Code

The fractionation of the genome into families of isochores allowed localizing sequences in the genome. The first localizations were those of mammalian globin genes. These showed that while the GC-poor β and γ globin genes were located in L2 isochores, the GC-rich α globin gene was located in H3 isochores, indicating that the GC levels of globin genes are correlated with the base composition of the isochores in which they are embedded [40]. The unavailability of additional gene probes at that time led us to concentrate on the genome localization (1) of two families of repeated sequences, LINEs and SINEs; and (2) of integrated viral sequences, since we had previously observed that the GC-rich bovine leukemia virus, BLV, was integrated in GC-rich H3 isochores [41].

The first approach led to the demonstration that GC-poor LINEs and GC-rich SINEs were mainly located in GC-poor isochores L1 and L2 [42] and GC-rich isochores H2 and H3 [43], respectively, another indication of a correlation between GC levels of specific sequences and GC levels of the “host” isochores. Since the two sets of repeats occupied vast regions, this also was an early evidence of a compositional bimodality in the human genome.

The second approach indicated that a compositionally matching integration was shown not only by BLV (as already mentioned), but also by all the GC-rich and GC-poor viral sequences tested. In fact, the viral sequences were expressed only when a compositionally matching integration was satisfied, namely when the situation was that of host genes [44,45].

Later on, an increasing number of protein-coding sequences were localized and indicated that linear compositional correlations were holding between exons (and their codon positions) and the isochores in which they were embedded [19,46,47], as well as between exons and the corresponding introns [26,48–50].

In fact, the correlations between coding and flanking non-coding sequences and those among the three codon positions led to a first definition of the “genomic code” [27,51], which stressed the fact that genomes are not random assemblies of coding and non-coding sequences, but systems obeying precise organization rules. This viewpoint was supported by: (1) the compositional compartmentalization of the human genome into isochores; (2) the localization of GC-poor/GC-rich isochores in Giemsa and Reverse bands, as well as their association with late and early replications; and (3) the evolutionary origin of GC-rich isochores and the bimodal structure of the genome of warm-blooded vertebrates (see the following section).

4. Gene Spaces

An important point concerning the distribution of coding sequences in the genome was its bimodality, as previously observed for GC-rich and GC-poor repeated sequences and for integrated viral sequences. Indeed, gene concentration was low and increased slowly with increasing GC in GC-poor isochores, whereas it increased rapidly in GC-rich isochores [29,36,45–47,52] reaching the highest level in the H3 isochore family, which was called the “genome core” [27,28], a definition later extended to all gene-rich isochores, including the H1 and H2 families. In fact, the break between the two slopes of gene concentration [47] defined two “gene spaces” [29], the GC-poor, gene-poor “empty quarter”, later called “genome desert” [45], and the GC-rich, gene-rich “genome core” already mentioned. As expected, the bimodality in protein-coding gene concentration could also be defined (see Figure 1B) using the 19,179 protein-coding genes of the bovine genome [37].

Another difference between the two gene spaces concerned the structures of the genes present in them. The first observation along this line concerned the GC-poor human dystrophin gene which was shown to be over 1 Mb in size [53] in sharp contrast with the housekeeping genes present in GC-rich isochores and tissue-specific genes present in GC-poor isochores, that are 10–50 Kb in size. This contrast was also seen in the large number of GC-poor and GC-rich genes explored recently (Bettecken, T., Moore, A. and

Bernardi, G. to be submitted), pointing to two structural classes of genes characterized by different encodings of amino acids and very different intron sizes.

The bimodality of gene distribution was followed by the discovery [38], confirmed and extended by Gilbert et al. (2004) [54], Federico et al. (2006) [55] and Dekker (2007) [56], that the genome core and the genome desert corresponded to open and closed chromatin and were centrally and peripherally located, respectively, in the interphase nucleus of vertebrates (Figure 1C; see Supplementary Table S1 for the properties of gene spaces). In fact, the open and closed chromatin corresponded to euchromatin and heterochromatin, two well-known structures from classical microscopy [57].

As already mentioned, at the evolutionary transition between cold- and warm-blooded vertebrates, the open chromatin of the genome core underwent a GC increase to be stabilized thermodynamically at the higher body temperature of warm-blooded vertebrates (higher GC levels were also observed in 18S RNAs from warm- vs. cold-blooded vertebrates [58]). In contrast, the closed chromatin of the GC-poor genome desert did not evolve towards higher GC levels (see the following section). These observations [45,59] stressed the existence of structural differences at the DNA level between the two gene spaces.

Gene spaces were also clearly detected by mapping DNase-I hypersensitive sites on human isochores [60]. Indeed, hypersensitive site regions (less protected by nucleosomes) are characterized by higher GC levels than the average GC level of the human genome (see Supplementary Figure S1A).

Remarkably, the early apoptotic chicken DNA fragmentation targets a number of specific open chromatin regions with high GC levels, also targeted by micrococcal nuclease digestion with, however, a lower level of GC specificity [61] (see Supplementary Figure S1B).

5. The Lamina- and Nucleolus-Associated Domains, Lads and Nads

A new approach, DamID, led to the discovery that genome–lamina interactions occur through more than 1300 sharply defined large domains 0.1–10 Mb in size corresponding to ~35% of the human genome, the Lamina-Associated Domains, LADs, typified by low gene expression levels and demarcated by the insulator protein CTCF, by promoters oriented away from LADs or by CpG islands [62]. In fact, LADs correspond in their properties to GC-poor isochores, that were previously localized at the periphery of the interphase nucleus [38] and also represent ~35% of human DNA. Later investigations [63] showed that LADs were characterized by low gene density, scarcity of CpG islands, high LINEs, low SINEs, and structural conservation across the cold- to warm-blooded vertebrate transition. In contrast, InterLADs showed opposite properties, corresponding to the features of TADs, the Topologically Associated Domains (see the following section, and also Supplementary Table S1). Nucleolus-associated domains, NADs, also exist and show a substantial functional overlap with LADs [64,65].

6. Spatial Compartments

Almost 20 years ago, a new approach, Capturing Chromatin Conformation, 3C, was developed to detect the frequency of physical interactions between any pair of genome *loci* [66,67]. The 3C approach was the starting point of an analysis of chromatin structure quite popular in recent years, the Hi-C approach [68], that probes the 3D architecture of the genome by coupling proximity-based ligation with massively parallel sequencing. This approach was claimed to have “identified an additional level of genome organization characterized by the spatial segregation of open, gene-rich, actively transcribed and closed, gene-poor, repressed chromatin to form two genome-wide spatial compartments” that were “arbitrarily” labelled A and B. This statement should, however, be corrected, since the two compartments correspond to the gene spaces (see Figure 1C) that, in fact, had been previously characterized in more detail (see Supplementary Table S1).

The Hi-C approach represented, however, a revolution in the study of chromatin in that (1) it was the first attempt to quantify the proximity of chromatin structures and to

use this criterion to define the compartments; (2) it led to the discovery of the short-range spatial organization of genomes into TADs [69–71] and of sub-compartments [72]. The Hi-C approach did not provide, however, any information on the DNA present in the compartments, on the evolutionary origins of the latter, nor on molecular explanations concerning the “folding principles of the human genome” [68], namely, the formation of compartments and of TADs, that were later provided by the genomic code.

7. Forests and Prairies

The high and low densities of CpG islands of the human genome were used to identify two genomic domains that were called “forests” and “prairies”, respectively [73]. According to the authors, this division partitions the genome into two types of regions that are genetically, epigenetically and transcriptionally different, and “outperform” isochores in the segregation of these properties. These conclusions [74,75] deserve two comments.

The first comment is that forests and prairies confirmed the bimodality of the genome as based on GC levels, gene concentration and DNA structure, and essentially coincide with the “gene spaces” (see Supplementary Table S1). This was expected since CpG islands increase in density with GC levels of isochores [76,77].

The second comment concerns the “outperformance” of forests and prairies compared to separations based on isochores. This point is, in fact, disputable for three reasons: (1) “CpG density distribution can vary greatly among species” [73] as previously found by Aissani and Bernardi, 1991 [76]), as well as within the human genome [78], which implies a lack of generality of the forests and prairies model in contrast with the case of isochores; (2) “Forests and prairies show enhanced segregation from each other in development, differentiation and senescence” [73], again implying a lack of generality; (3) “Genomes having high and uniform CpG distribution can be considered as consisting of mainly forests with little mosaicity”; this conclusion is incorrect as shown by the fact that one of these genomes, the *Drosophila* genome, shows (like some other invertebrate genomes) both high CpG levels and isochores that clearly belong in three families [79].

8. Genome Compartments and Isochore “Super-Families”

A new approach to the study of isochores, a GC profiling of 100 Kb blocks [6], showed that the isochores of the human genome group are placed into two “super-families” characterized by two long-range 3D structures [8] (see Figure 3A,B and Figure 2A): (1) the GC-poor, low compositional heterogeneity isochores of the L1 family and of the L2⁻ sub-family, both of which correspond to LADs (as well as to the “genome desert”); and (2) the GC-rich isochores that comprise the single- and the multi-peak isochores of the H1–H3 families, and of the L2⁺ sub-family, that correspond to InterLADs, and to the “genome core”.

In human chromosome 21, the first, compositionally flat structure is found in regions 2, X and Y. In contrast, the second one comprises the single peaks and the sets of peaks and valleys of regions 1, 3, 4, 5, 6 (see Figure 3A,B; similar situations are present in all human chromosomes; Jabbari, K., Ritucci, M., Bernardi, G., to be submitted). That the two compartments are “spatially arranged in a polarized manner” in chromosomes was also independently shown by Wang et al. (2016) [80]. Figure 3C shows that the two “super-families” of isochores correspond to the “spatial compartments”, as expected from the comparison of Figure 1C.

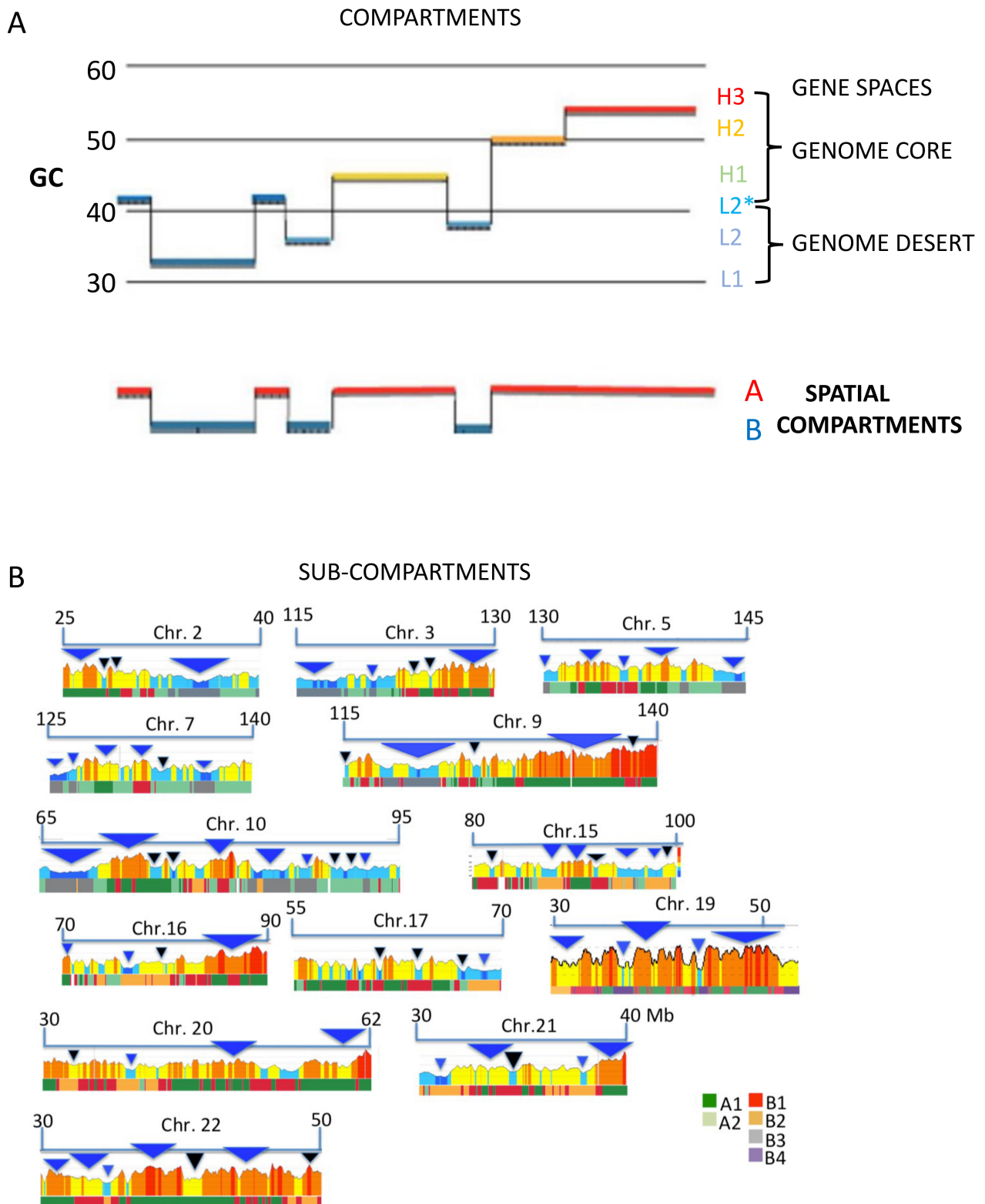


Figure 2. (A) A profile of a chromosome segment in which all isochores families are represented is compared with the corresponding compartment profile to show that, while the L1 and L2⁻ isochores correspond to the B compartment, the other isochores correspond to the A compartments (modified from [6]). (B) The figure shows an average match between isochores/isochores-blocks (profiles) and chromatin sub-compartments (A1–B4). Blue/black triangles are examples of good/bad matches (from a Supplementary Figure of [30]).

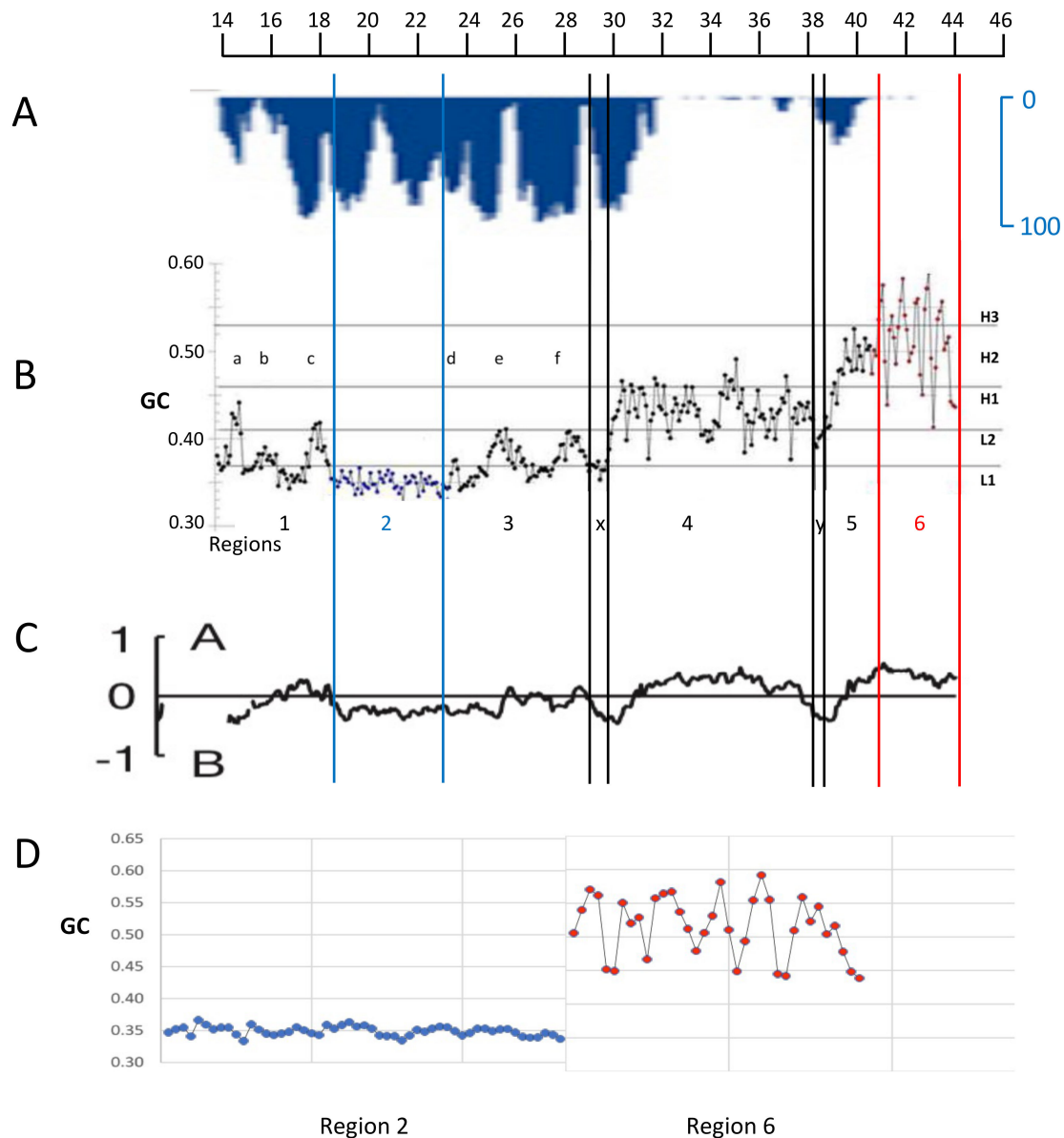


Figure 3. (A) The inverted LAD profile of the long arm of human chromosome 21 (from [81]) is compared with the 100 Kb dot-plot GC profile of Figure 2B (from [7]). LADs correspond to L1 isochores (two blue lines bracket region 2, the largest L1 isochore) and to two “valley” isochores (double black lines, X and Y, that belong in the L2⁻ sub-family, a low heterogeneity subfamily in the L2 isochore GC range). One H1 (a), and five L2⁺ single peak isochores (b–f) correspond to InterLADs (the L2⁺ sub-family is a heterogeneous sub-family in the L2 isochore GC range). (C) The 100 Kb profile is compared with the A and B compartments (adapted from [82]; the thickness of the profile is due to the enlargement of the original minute figure). Imperfect alignments of Figures (A–C) are due to the different original sources of the panels. (D) The enlarged GC profiles of regions 2 and 6 (corresponding to the telomere) show the striking differences in their compositional ranges (see text).

Finally, Figure 3D presents an enlargement of two typical regions, 2 and 6, better showing the sharp structural contrast between them, the former covering a ~3% GC range centered at ~35% GC, the latter a ~15% GC range with a lowest level at ~45% GC. In conclusion, the results of Figure 3 not only establish a precise link between the 3D DNA structures of isochore “super-families” and the genome compartments, but also show the existence of different 3D structures in the two different isochore “super-families”.

9. Genome Sub-Compartments and Isochores

Several lines of evidence indicated the existence of finer correlations of isochores (1) with the “sub-compartments” A1, A2 and B1–B3 based on Hi-C [72] that showed links with gene expression and histone modifications; in this case, A1 sub-compartments were shown to correspond to H2/H3 isochores, A2 sub-compartments to H1 and L2⁺ isochores, and B1–B3 sub-compartments to L2⁻ and L1 isochores (see Figure 2B [30]); (2) with LADs and TADs (see Figure 4A), a result obtained for all human and mouse chromosomes [30]; (3) with loop domains (see Figure 4B) [30,83].

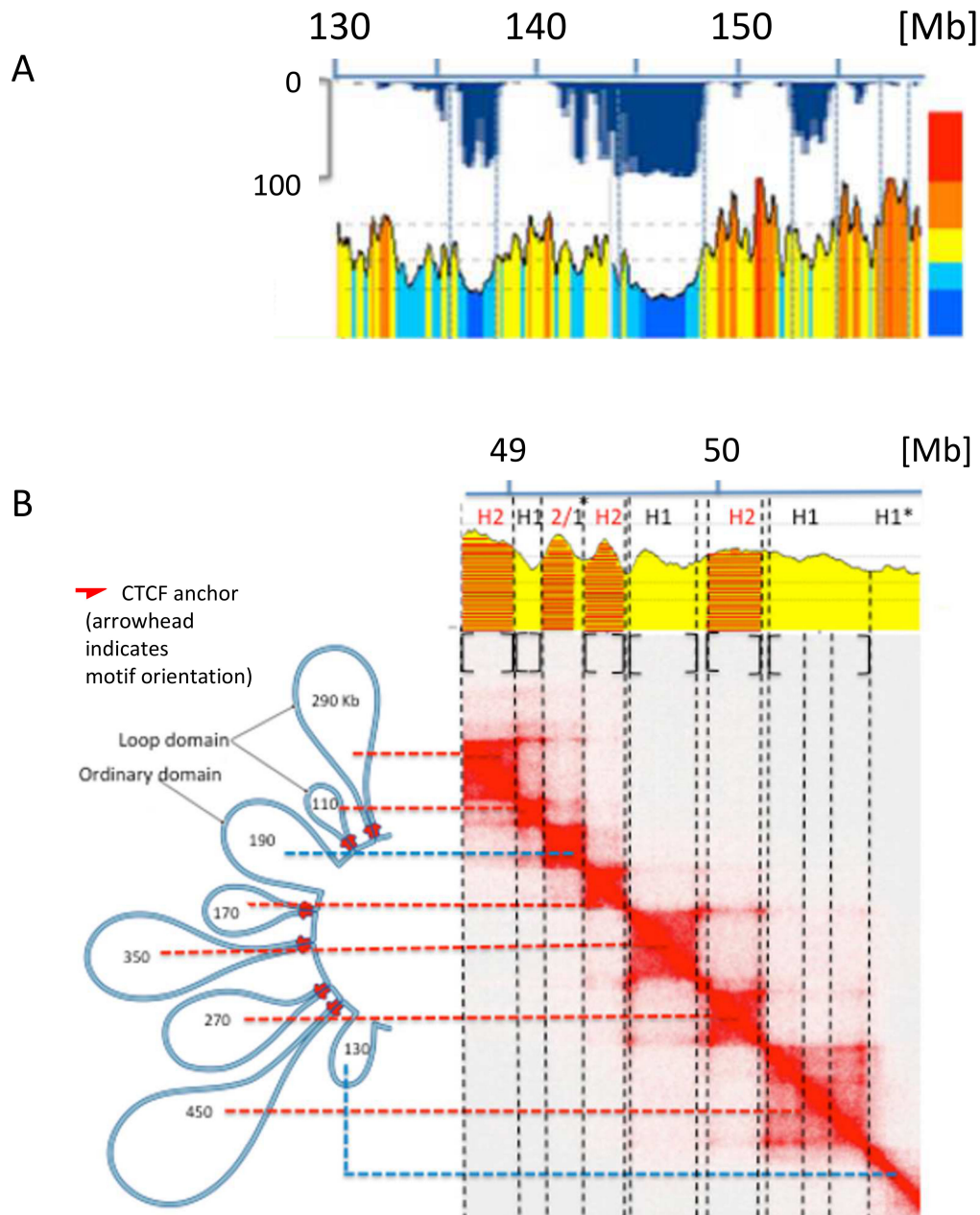


Figure 4. (A) ~30 Mb of human chromosome 7 is analyzed at a resolution of 50 Kb; a fine correspondence of isochores boundaries with LAD and TAD boundaries can be observed (modified from [30]). (B) The chromatin loops from a 2.1 Mb region of human chromosome 20 have been aligned with the corresponding heat map which was used to segment the corresponding DNA sequence into isochores; asterisks indicate small anomalies in the isochores/domains correspondence (modified from [30]).

In contrast with the compositionally flat B1–B3 sub-compartments, corresponding to L2⁻ and L1 isochore super families, the A1 and A2 sub-compartments corresponding to L2⁺, H1–H3 isochores are characterized by sharp peaks. In this case, compositional gradients (involving oligo-G’s, CpG’s and CpG islands) characterized by increasing bendability, increasing accessibility, decreasing supercoiling and decreasing nucleosome density [84], constrain chromatin into loops, the tips of the loops corresponding to the highest GC levels and being the attachment sites for cohesin, while the increasing presence of oligo-A’s in the valleys helps the folding of the chromatin fibre between GC-rich peaks. This initial “moulding step” leads to cohesin-free “primary TADs” [7,8] that may be followed by extrusion through the action of cohesin, the extrusion being stopped by GC-poor DNA-bound CTCF (see Figure 5). At this point, it should be stressed that the results obtained on a DNA basis and those obtained by Hi-C are linked with each other, as expected.

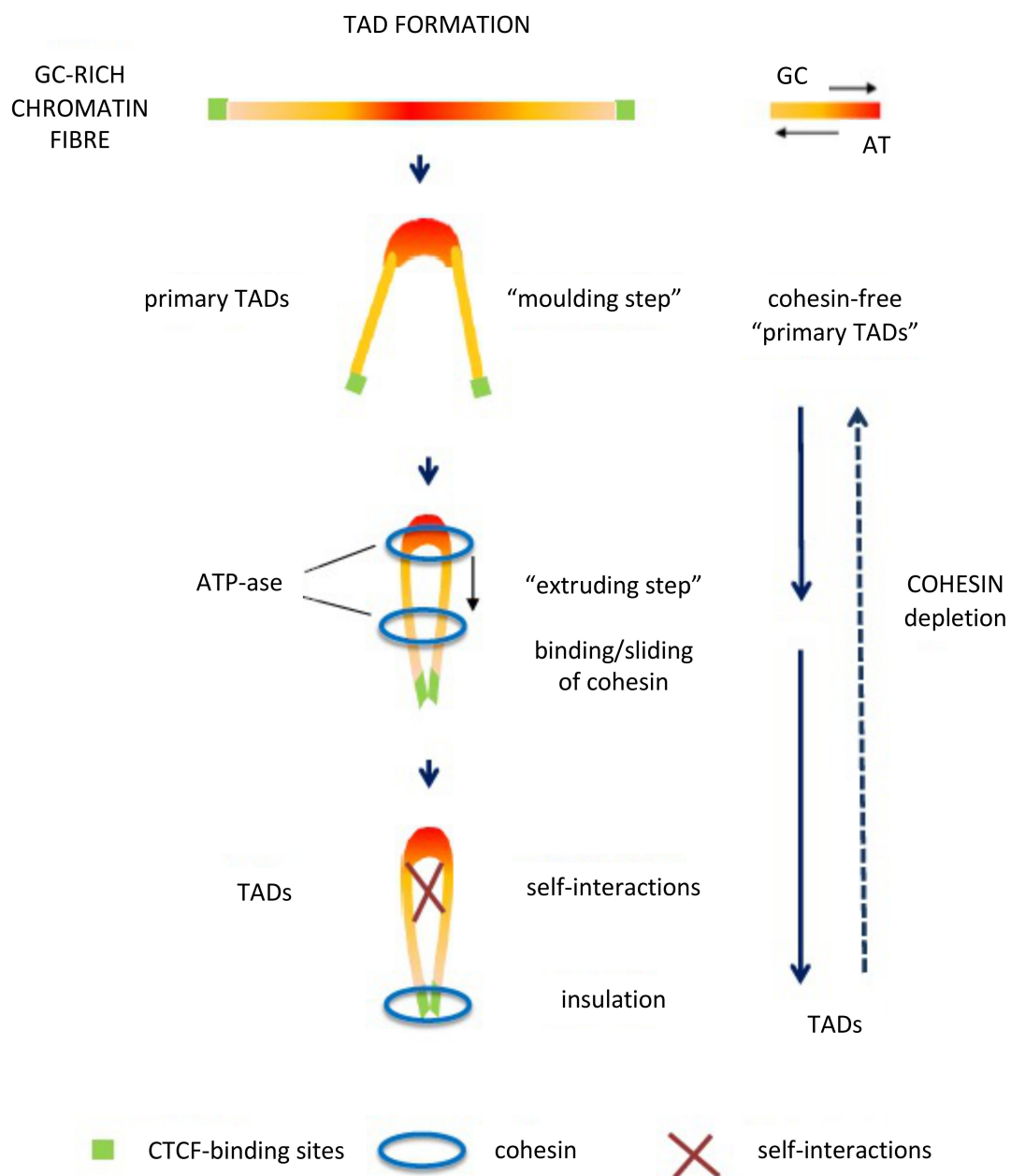


Figure 5. The proposed model for the formation of “primary TADs” and TADs. (Modified from [8]; See text).

Extrusion has, however, several problems: (1) removal of cohesin and CTCF from chromosomes showed limited effects on steady-state transcription [85–87]; (2) extrusion does not appear to be a general phenomenon in chromatin domains; indeed, it is “malleable and variable between individual cells” [88] and absent in *Drosophila* [89] (in which case building of TADs does not appear to make use of the CTCF/cohesin loop extrusion mechanism) and in many other cases; for instance, the Hi-C approach can fail to detect known structures such as interactions with nuclear bodies, because these DNA regions can be too far apart to directly ligate [90]; (3) extrusion can take place not only through the classical two-sided manner but also through a one-sided manner raising another unsolved problem [91].

Sub-compartments were very recently analyzed by SPIN [92], i.e., by combining compartment mapping and chromatin interaction data. This approach clarified the patterns of specific sub-compartments relative to nuclear speckles, lamina and nucleolus and found specific sequence repeats in sub-compartments.

Another approach to study 3D genome organization and evolution was proposed by Mourad (2019) [93] on the basis of genome sequences only, more specifically by comparing the distances between convergent and divergent CTCF motifs (ratio R) to detect CTCF looping. R values were very high for L1 isochores and sub-compartments B3 (corresponding to heterochromatin) and very low for H3 isochores.

10. Short Sequences in Isochores and Nucleosome Binding

Specific short sequences are characteristic of isochore families, as shown by the “short sequence designs” of isochores [15,94] and the following results: (1) Previous investigations [95,96] indicated that specific sets of di- and tri-nucleotides, such as “AAA/TTT” and “GGG/CCC”, as well as the “A/T-only” and “G/C-only” trinucleotides, have widely different distributions in different isochore families of the human genome, the former being predominant in GC-poor, the latter in GC-rich isochores, as expected; (2) Profiles of all 64 trinucleotides of chromosome 21 confirm an essentially bimodal topology with compositionally flat (yet different) levels in GC-poor isochores and single or multiple peaks in GC-rich isochores [9]; (3) Figure 6A,B shows the profiles of oligo-A’s and oligo-G’s of regions 2 and 6 of chromosome 21 that are typical of GC-poor and GC-rich isochores; these results indicate that region 2 shows essentially flat levels for 3G and 5G with small blocks of oligo A’s, whereas region 6 shows sharp peaks of both 3A and 5A, as well as of 3G and 5G.

In conclusion, two different classes of isochores can be distinguished on the basis of short-sequence profiles, an important point because these profiles define the nucleosome binding pattern. Indeed, investigations on DNA/nucleosome interactions have shown (1) that DNA sequences differ greatly in their ability to bend sharply [97,98] and that nucleosome formation is highly dependent on specific DNA sequences [99,100], AT and GC tracts being intrinsically inhibitory [101,102]; and (2) that CpGs and CpG islands that increase in density with GC levels are strongly inhibitory for nucleosome binding [102–104]. These observations led to the concept of a “nucleosome positioning code”, interestingly also called “a genomic code for nucleosome positioning” [101].

An important new point is that nucleosome positioning is bimodal, in that it is denser in the regions corresponding to GC-poor isochores that are characterized by the compositionally least heterogeneous L1 and L2[−] isochores. In this case, larger and deeper “clutches” of nucleosomes correspond to the “closed” heterochromatin, whereas “open chromatin” is formed by smaller and less dense clutches that associate with RNA Polymerase II [105]. Indeed, the internucleosome sequences are increasingly broader with increasing GC and increasing amounts of sequences (like CpGs and CpG islands) that are incompatible with nucleosome binding. Finally, there is evidence that nucleosomes remember where they were [106]. In other words, there are two sets of nucleosome spacings, a short one and a long one (see Figure 6C and Figure S1A,B).

Interestingly, the trinucleotide patterns in the 0.1–0.5 Kb sequences flanking genes located in GC-rich and GC-poor isochores indicate differences in the transcription factors

that bind upstream and downstream of genes [96]. This result indicates differences in the regulation of genes located in different isochore families, in agreement with the fact that different classes of genes are located in different isochore families [107].

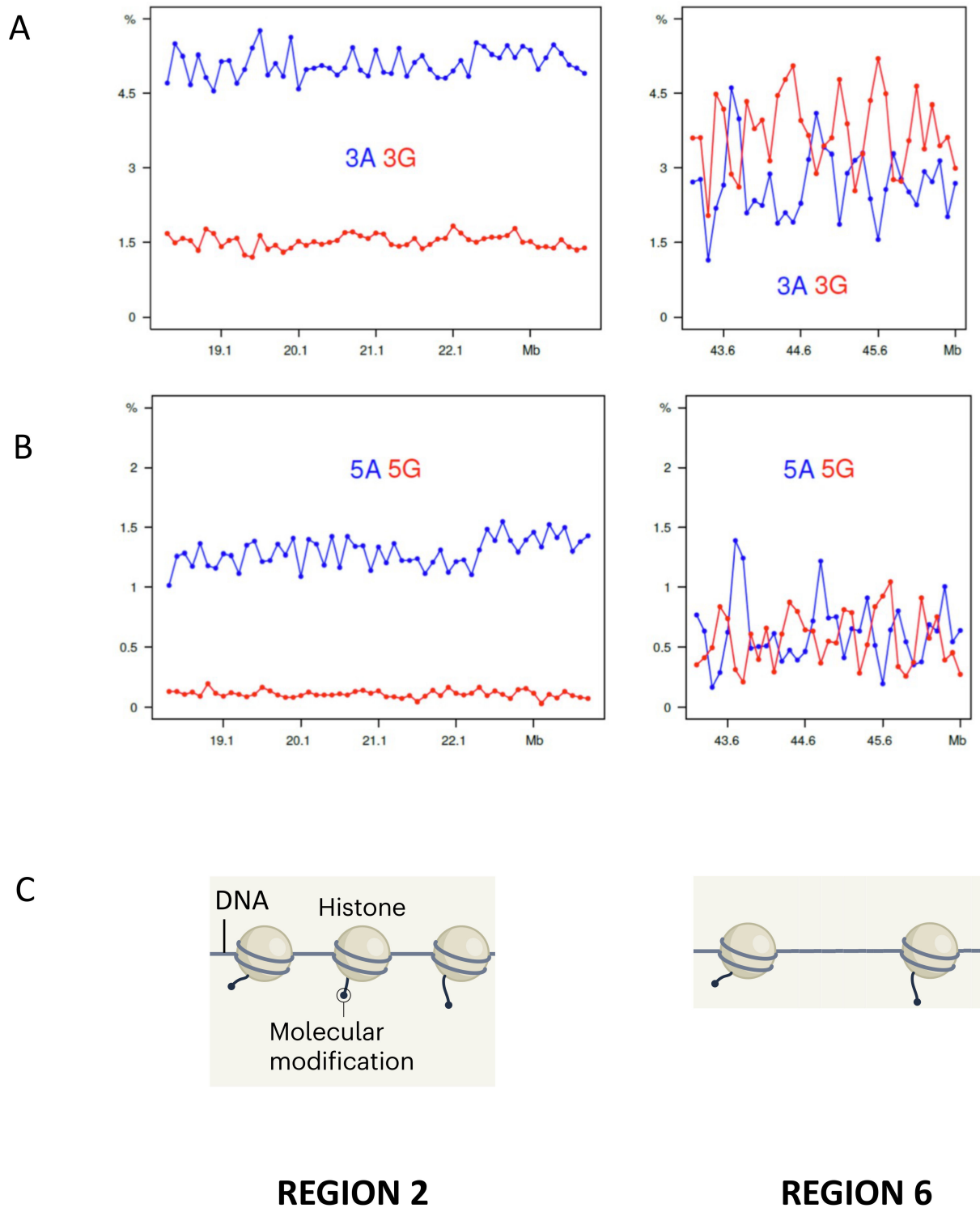


Figure 6. (A,B) Profiles of tri- and penta- A's and G's of regions 2 and 6 (from Figure S3 of [108]) (C) A scheme of nucleosome densities for regions 2 and 6. (Modified from [109]; See also legend of Figure 7).

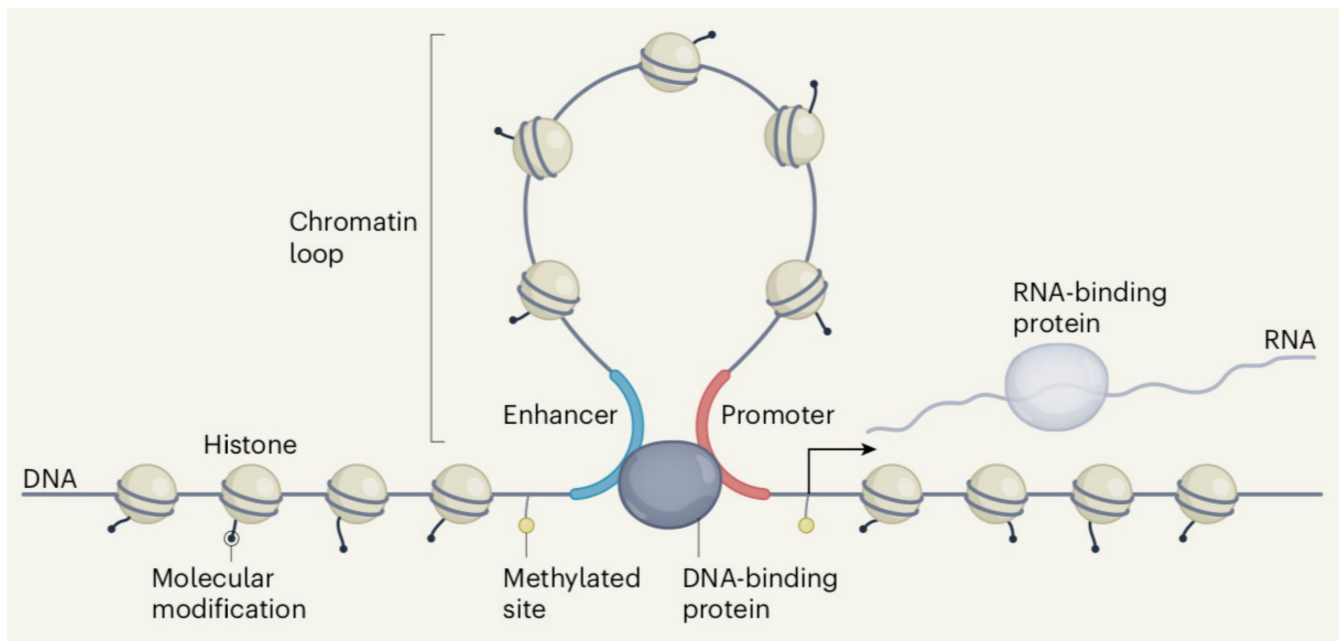


Figure 7. Functional elements across the human genome [109]. This Figure can also be seen as a model representing the structural elements discussed in the article (see text). This figure is published with the permission of authors and Springer Nature.

11. Conclusions

11.1. The Correlations of DNA Sequences with Chromatin Architecture

These correlations were investigated at three DNA size levels (see Supplementary Table S2). The long-range bimodal compartmentalization of chromatin was shown to be associated with the different DNA structures of the two “super-families” of isochores, a compositionally flat one, and a single peak or peak-and-valley one. The demonstration of correlations between the 3D structure of isochore “super-families” and compartments is important, if one considers that it is currently admitted that the “exact molecular players that drive compartment organization are not yet known” [108] and that “the molecular determinants that modulate the maintenance and movement of compartmentalization remain largely elusive” [92] or are due to “self-associating homeotypic chromatin types” [88]. In fact, “attractions between heterochromatin regions were proposed to be essential for the phase separation of the active and inactive genomes”, and “interactions of the chromatin with the lamina are necessary to build the conventional architecture from these segregated phases” [110] (Falk et al. 2019). The interaction of GC-poor sequences from both cold- and warm-blooded vertebrates with lamina provides support to this proposal.

At a short-range resolution, “primary TADs” [7–9] appear to be constrained into loops by the properties that accompany the GC gradients of isochore peaks, to which they correspond, namely, increasing bendability, increasing nuclease accessibility, decreasing supercoiling and decreasing nucleosome density. This is again an important conclusion, if one considers that extrusion was visualized as the only way to explain the formation of the loops [108,111], a view also contradicted by the fact that extrusion may or may not affect “primary TADs”. Two important results concern the correspondence of isochores and sub-compartments with LADs and TADs (Figure 4A) and with loop domains (Figure 4B). The compositional stability of the GC-poor isochores of LADs across the evolutionary transition between cold- and warm-blooded vertebrates are most likely to be due to their interactions with the lamina as already mentioned. In contrast, the isochores that are not interacting with lamina underwent a GC increase (and an increase in compositional heterogeneity) in order to be stabilized at the higher body temperature of warm-blooded vertebrates. Needless to say, the correlations of DNA sequences with genome organization

explain why isochores are “a fundamental level of genome organization” [112], even if when this was proposed on the basis of our work, it was not clear what the real reason was.

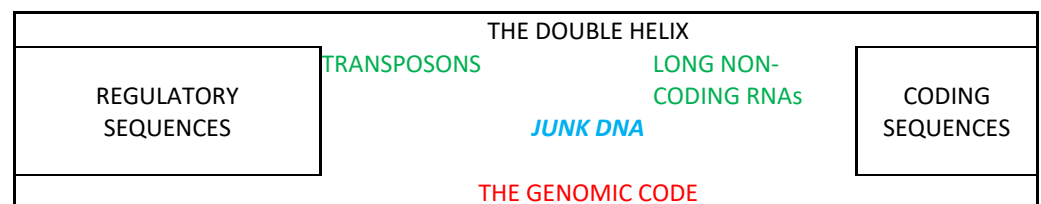
Finally, at the short sequence level, the two sets of sequences, GC-rich and GC-poor, that characterize the compartments and sub-compartments define nucleosome positioning. It is understandable that the thermal transition is accompanied by a decreased density of nucleosomes, because of the increasing amounts of GC-rich sequences (such as CpG and CpG islands) that are incompatible with nucleosome binding and that are located in the internucleosome linkers.

It should be stressed that the structure of DNA sequences and not nucleosome distribution is primarily responsible for moulding chromatin architecture. Indeed, DNA structure not only exists independently of nucleosome binding, but it is responsible for it. This is shown by the striking finding that TADs are not very different in fibroblasts and in spermatozoa [113], in which latter case nucleosomes are replaced by protamines (see Supplementary Figure S2).

11.2. The Genomic Code

The properties of DNA sequences, as seen at the levels just discussed, mould chromatin architectures according to a “genomic code” [6–8] coding sequences and the regulatory sequences (see Table 1). The importance of the sequences under discussion is underlined by the fact that they pervade the totality of the genome (including satellite DNAs; see Figure S3), overlapping and compositionally constraining not only transposons and long non-coding RNAs, but also the coding sequences. Indeed, the composition of codons for amino acids varies according to the composition of the corresponding isochores.

Table 1. The four pillars of the genome ^(a).



^(a) Titles in the diagram: Black: The original three pillars. The double helix fills the totality of the genome DNA. The coding sequences correspond to less than 2% of the genome. The regulatory sequences also correspond to a minute part of the genome. Blue: The “junk DNA” fills in the space not occupied by the coding and regulatory sequences. Its existence is ruled out by the genomic code. Green: Transposons and long non-coding RNAs. Red: The genomic code, the fourth pillar, is pervasive and fills the totality of the genome.

The genomic code defines two vast ensembles in the vertebrate genome. The first is characterized by the GC-poorest, least heterogeneous sequences and represents ~35% of the genome, that are associated with the lamina, have a high nucleosome density, and are compositionally stable through the cold- to warm-blooded vertebrates transition. The second one, corresponding to the GC-rich sequences, ~65% of the genome, is characterized by increasing compositional heterogeneity with increasing number of sequences that are incompatible with nucleosome binding, leading, as a consequence, to a wider internucleosome spacing.

It should be stressed that the approach used in the experiments presented, based as it is on the role of DNA sequences in the moulding of chromatin architecture, is totally different from the strategy prevailing over the last few years in the field of chromatin, a strategy which is based on the proximity of chromatin structures. It is not surprising, therefore, that while the latter strategy discovered sub-compartments and TADs, it could not provide explanations at the molecular level for the “folding principles of the human genome” [68] that could be obtained by using DNA sequences.

One can now ask whether any of the properties of the genomic code were reported before: (1) Trifonov (1980) [114] remarkably predicted the sequence-dependent deformational anisotropy of chromatin DNA, as well as the possibility of overlapping codes,

one of which, the “chromatin code”, would provide instructions on appropriate placement of nucleosomes along DNA molecules and their spatial arrangement [115]; (2) the work of Todolli et al. (2017) [116] pointed to local sequence-dependent features found in high-resolution structures that introduce irregularities in the disposition of adjacent residues that facilitate the specific binding of proteins and also determine the positions of nucleosomes on DNA and the lengths of the interspersed DNA linkers; (3) Ramirez et al. (2018) [117] identified eight DNA motifs enriched at TAD boundaries and proposed that DNA sequence guides the genome architecture by allocation of boundary proteins in the genome; (4) Gorkin et al. (2018) [118] identified thousands of regions in the lymphoblastoid cell lines from 20 individuals where 3D chromatin varies and demonstrated that common DNA sequence variants can influence chromatin conformation, a result expected from the moulding of chromatin by DNA sequences [6–9]; (5) Fudenberg et al. (2020) [119] and Schwesinger et al. (2020) [120] developed methods that predict genome folding from DNA sequences alone. In other words, the old predictions of Trifonov and recent results are in agreement with the present conclusions.

At this point, three remarks should be done. (1) Satellite DNA sequences (see Supplementary Figure S3 for an example) appear at precise locations; (2) Non-B-form DNA is enriched at centromeres [121]; (3) Transposons and long-non-coding RNAs are expressed while overlapping DNA.

11.3. The Genomic Code and “Junk DNA”

When it was realized that the coding sequences represented a small percentage of the human genome, it was proposed that the majority of it was “junk DNA” [5]. This was the beginning of the longest debate in molecular evolution, which is still going on after almost 50 years [122–126], justified as it was by the lack of alternative explanations. The discovery of the moulding of chromatin structure by DNA sequences, the genomic code, leads in fact to an explanation (see Table 1) which is alternative to the idea that “perhaps 90% of our DNA, though biochemically active, does not contribute to fitness in any sequence-dependent way and possibly in no way at all” [126]. Indeed, the sequences involved in moulding chromatin (and contributing to fitness) occupy the totality of genome DNA leaving no room for “junk DNA”. It should be stressed, however, that while single nucleotide changes in the sequences that are the backbones of chromatin are unlikely to lead to structural variations being neutral [127,128] or nearly neutral [129,130], short-sequence deletions/insertions, can lead to somatic structural variations such as the fusion of TADs and complex rearrangements that markedly change chromatin folding maps in the cancer genomes [131] and most probably in aging and a number of diseases.

11.4. The Genomic Code and ENCODE

At this point, it is of interest to compare the views of the human genome as derived from the structural approach discussed so far and from ENCODE, the Encyclopedia of DNA Elements, which is focused on the functional properties of the genome. Figure 7 presents the functional elements as derived from ENCODE [109]. This concept is of great interest because it can be also read as representing the picture derived from the structural work discussed here. Indeed, there are two major structural elements, described so far, that appear in the figure. The horizontal bottom line of nucleosomes can be seen as corresponding to LADs, whereas the loop corresponds to a TAD. (Incidentally, the nucleosome spacing correctly happens to be larger in the latter case compared to the former one). In other words, the Figure can be read not only as presenting the functional elements but also the structural elements. In other words, the results of our structural approach can be seen as matching the results of the functional approach of ENCODE.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/life11040342/s1>, Table S1: Human Genome Compartmentalization, Table S2: The Bimodal Organization of The Human Genome at Three DNA Size Levels, Figure S1: A. Compositional distribution of human DNA and of hypersensitive sites, Figure S2: TADs as present in sperm cells and fibroblasts of a region of chromosome 19, Figure S3: Compositional profile of the short arm and of the centromere (α -satellite) of chromosome 21 (by Gregorio Bernardi).

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This article is dedicated to the memories of Hugo Theorell (1905–1982) in whose laboratory I started my MD thesis (in the field of Biochemistry) and of Jacques Monod (1910–1976), the Honorary President of the Jury of my thesis in Physics. The author thanks Paolo Ascenzi for hospitality, Giacomo Bernardi for discussions, Gregorio Bernardi for Supplementary Figure S3, Salvo Saccone for Figure 1C, and Marta Ritucci for excellent technical help.

Conflicts of Interest: The author declares no competing interests.

References

- Winkler, H. *Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreich*; Fischer: Jena, Germany, 1920.
- Watson, J.D.; Crick, F.H. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171*, 737–738. [[CrossRef](#)]
- Jacob, F.; Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **1961**, *3*, 318–356. [[CrossRef](#)]
- Nirenberg, M.W.; Matthaei, J.H. The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1588–1602. [[CrossRef](#)] [[PubMed](#)]
- Ohno, S. So much “junk” DNA in our genome. *Brookhaven Symp. Biol.* **1972**, *23*, 366–370. [[PubMed](#)]
- Bernardi, G. The genomic code: Isochores encode and mold chromatin domains. *BioRxiv* **2016**, 096487.
- Bernardi, G. The formation of chromatin domains involves a primary step based on the 3-D structure of DNA. *Sci. Rep.* **2018**, *8*, 17821. [[CrossRef](#)] [[PubMed](#)]
- Bernardi, G. The Genomic Code: A Pervasive Encoding/Molding of Chromatin Structures and a Solution of the “Non-Coding DNA” Mystery. *BioEssays* 2019. [[CrossRef](#)]
- Lamolle, G.; Sabbia, V.; Musto, H.; Bernardi, G. The short-sequence design of DNA and its involvement in the 3-D structure of the genome. *Sci. Rep.* **2018**, *8*, 17820–17828. [[CrossRef](#)]
- Bernardi, G.; Champagne, M.; Sadron, C. Kinetics of the Enzymatic degradation of deoxyribonucleic acid into sub-units. *Nature* **1960**, *188*, 228–229. [[CrossRef](#)]
- Filipinski, J.; Thiery, J.P.; Bernardi, G. An analysis of the bovine genome by $\text{Cs}_2\text{SO}_4/\text{Ag}^+$ density gradient centrifugation. *J. Mol. Biol.* **1973**, *80*, 177–197. [[CrossRef](#)]
- Tapiero, H.; Caneva, R.; Schildkraut, C.L. Fractions of Chinese hamster DNA differing in their content of guanine + cytosine and evidence for the presence of single-stranded DNA. *Biochim. Biophys. Acta* **1972**, *272*, 350–360. [[CrossRef](#)]
- Thiery, J.P.; Macaya, G.; Bernardi, G. An analysis of eukaryotic genomes by density gradient centrifugation. *J. Mol. Biol.* **1976**, *108*, 219–235. [[CrossRef](#)]
- Macaya, G.; Thiery, J.P.; Bernardi, G. An approach to the organization of eukaryotic genomes at a macromolecular level. *J. Mol. Biol.* **1976**, *108*, 237–254. [[CrossRef](#)]
- Cuny, G.; Soriano, P.; Macaya, G.; Bernardi, G. The major components of the mouse and human genomes. 1. Preparation, basic properties and compositional heterogeneity. *Eur. J. Biochem.* **1981**, *115*, 227–233. [[CrossRef](#)]
- Costantini, M.; Clay, O.; Auletta, F.; Bernardi, G. An isochore map of human chromosomes. *Genome Res.* **2006**, *16*, 536–541. [[CrossRef](#)]
- Liu, F.; Tøstesen, E.; Sundet, J.K.; Jenssen, T.K.; Bock, C.; Jerstad, G.I.; Thilly, W.G.; Hovig, E. The human genomic melting map. *PLoS Comp. Biol.* **2007**, *3*, 874. [[CrossRef](#)]
- Cozzi, P.; Milanese, L.; Bernardi, G. Segmenting the human genome into isochores. *Evol. Bioinform.* **2015**, *11*, 253–261. [[CrossRef](#)]
- Bernardi, G.; Olofsson, B.; Filipinski, J.; Zerial, M.; Salinas, J.; Cuny, G.; Meunier-Rotival, M.; Rodier, F. The mosaic genome of warm-blooded vertebrates. *Science* **1985**, *228*, 953–957. [[CrossRef](#)]
- Saccone, S.; De Sario, A.; Wiegant, J.; Raap, A.K.; Della Valle, G.; Bernardi, G. Correlations between isochores and chromosomal bands in the human genome. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11929–11933. [[CrossRef](#)]
- Costantini, M.; Bernardi, G. Replication timing, chromosomal bands and isochores. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 3433–3437. [[CrossRef](#)]
- Bernardi, G. Chromosome Architecture and Genome Organization. *PLoS ONE* **2017**, *10*, e0143739.

23. Federico, C.; Saccone, S.; Bernardi, G. The gene-richest bands of human chromosomes replicate at the onset of the S-phase. *Cytogenet. Cell Genet.* **1998**, *80*, 83–88. [[CrossRef](#)] [[PubMed](#)]
24. Watanabe, Y.; Fujiyama, A.; Ichiba, Y.; Hattori, M.; Yada, T.; Sakaki, Y.; Ikemura, T. Chromosome-wide assessment of replication timing for human chromosomes 1 q and 21q: Disease-related genes in timing-switch regions. *Hum. Mol. Gen.* **2002**, *11*, 13–21. [[CrossRef](#)] [[PubMed](#)]
25. Schmegner, C.; Hameister, H.; Vogel, W.; Assum, G. Isochores and replication time zones: A perfect match. *Cytogenet. Genome Res.* **2007**, *116*, 167–172. [[CrossRef](#)]
26. Bernardi, G.; Bernardi, G. Compositional constraints and genome evolution. *J. Mol. Evol.* **1986**, *24*, 1–11. [[CrossRef](#)]
27. Bernardi, G. The vertebrate genome: Isochores and evolution. *Mol. Biol. Evol.* **1993**, *10*, 186–204.
28. Bernardi, G. The human genome: Organization and evolutionary history. *Annu. Rev. Genet.* **1995**, *29*, 445–476. [[CrossRef](#)]
29. Bernardi, G. Isochores and the evolutionary genomics of vertebrates. *Gene* **2000**, *241*, 3–17. [[CrossRef](#)]
30. Jabbari, K.; Bernardi, G. An Isochore Framework Underlies Chromatin Architecture. *PLoS ONE* **2017**, *12*, e0168023. [[CrossRef](#)]
31. Hughes, S.; Zelus, D.; Mouchiroud, D. Warm-blooded isochore structure in Nile crocodile and turtle. *Mol. Biol. Evol.* **1999**, *16*, 1521–1527. [[CrossRef](#)]
32. Chojnowski, J.L.; Franklin, J.; Katsu, Y.; Iguchi, T.; Guillette, L.J., Jr.; Kimball, R.T.; Braun, E.L. Patterns of Vertebrate Isochore Evolution Revealed by Comparison of Expressed Mammalian, Avian, and Crocodylian. *Genes J. Mol. Evol.* **2007**, *65*, 259–266. [[CrossRef](#)]
33. Fortes, G.G.; Bouza, C.; Martinez, P.; Sanchez, L. Diversity in isochore structure among cold-blooded vertebrates based on GC content of coding and non-coding sequences. *Genetica* **2007**, *129*, 281–289. [[CrossRef](#)]
34. Chojnowski, J.L.; Braun, E.L. Turtle isochore structure is intermediate between amphibians and other amniotes. *Integr. Comp. Biol.* **2008**, *48*, 454–462. [[CrossRef](#)]
35. International Human Genome Sequencing Consortium. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409*, 860–921. [[CrossRef](#)]
36. Bernardi, G. Misunderstandings about isochores. Part 1. *Gene* **2001**, *276*, 3–13. [[CrossRef](#)]
37. Arhondakis, S.; Milanesi, M.; Castrignanò, T.; Gioiosa, S.; Valentini, A.; Chillemi, G. Evidence of distinct gene functional patterns in GC-poor and GC-rich isochores in *Bos taurus*. *Anim. Genet.* **2020**, *51*, 358–368. [[CrossRef](#)]
38. Saccone, S.; Federico, C.; Andreozzi, L.; D’Antoni, S.; Bernardi, G. Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. *Gene* **2002**, *300*, 169–178. [[CrossRef](#)]
39. Stevens, T.J. 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **2017**, *544*, 59–64. [[CrossRef](#)]
40. Bernardi, G. Organization and evolution of the eukaryotic genome. In *Recombinant DNA and Genetic Experimentation*; Morgan, J., Whelan, W.J., Eds.; Pergamon Press: New York, NY, USA, 1979; pp. 15–20.
41. Kettman, R.; Meunier-Rotival, M.; Cortadas, J.; Cuny, G.; Ghysdael, J.; Mammerickx, M.; Burny, A.; Bernardi, G. Integration of bovine leukemia virus DNA in the bovine genome. *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 4822–4826. [[CrossRef](#)]
42. Meunier-Rotival, M.; Soriano, P.; Cuny, G.; Strauss, F.; Bernardi, G. Sequence organization and genomic distribution of the major family of interspersed repeats of mouse DNA. *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 355–359. [[CrossRef](#)]
43. Soriano, P.; Meunier-Rotival, M.; Bernardi, G. The distribution of interspersed repeats is nonuniform and conserved in the mouse and human genomes. *Proc. Natl. Acad. Sci. USA* **1983**, *80*, 1816–1820. [[CrossRef](#)] [[PubMed](#)]
44. Rynditch, A.V.; Zoubak, S.; Tsyba, L.; Tryapitsina-Guley, N.; Bernardi, G. The regional integration of retroviral sequences into the mosaic genomes of mammals. *Gene* **1998**, *222*, 1–16. [[CrossRef](#)]
45. Bernardi, G. *Structural and Evolutionary Genomics: Natural Selection in Genome Evolution*; Elsevier: Amsterdam, The Netherlands, 2004; Available online: www.giorgiobernardi.eu (accessed on 12 April 2021).
46. Mouchiroud, D.; D’Onofrio, G.; Aissani, B.; Macaya, G.; Gautier, C.; Bernardi, G. The distribution of genes in the human genome. *Gene* **1991**, *100*, 181–187. [[CrossRef](#)]
47. Zoubak, S.; Clay, O.; Bernardi, G. The gene distribution of the human genome. *Gene* **1996**, *174*, 95–102. [[CrossRef](#)]
48. Ikemura, T.; Aota, S. Alternative chromatic structure at CpG islands and quinacrine-brightness of human chromosomes. Global variation in G+C content along vertebrate genome DNA. Possible correlation with chromosome band structures. *J. Mol. Biol.* **1988**, *60*, 909–920.
49. Bernardi, G. The isochore organization of the human genome. *Annu. Rev. Genet.* **1989**, *23*, 637–661. [[CrossRef](#)]
50. Costantini, M.; Bernardi, G. Correlations between coding and contiguous non-coding sequences in isochore families from vertebrate genomes. *Gene* **2008**, *410*, 241–248. [[CrossRef](#)]
51. Bernardi, G. Le Génome des Vertébrés: Organisation, Fonction et Evolution. *Biofutur* **1990**, *94*, 43–46.
52. Bernardi, G. The neo-selectionist theory of genome evolution. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 8385–8390. [[CrossRef](#)]
53. Bettecken, T.; Aissani, B.; Müller, C.R.; Bernardi, G. Compositional mapping of the human dystrophin-encoding gene. *Gene* **1992**, *122*, 329–335. [[CrossRef](#)]
54. Gilbert, N.; Boyle, S.; Fiegler, H.; Woodfine, K.; Carter, N.P.; Bickmore, W.A. Chromatin architecture of the human genome: Gene-rich domains are enriched in open chromatin fibers. *Cell* **2004**, *118*, 555–566. [[CrossRef](#)]
55. Federico, C.; Scavo, C.; Cantarella, C.D.; Motta, S.; Saccone, S.; Bernardi, G. Gene-rich and gene-poor chromosomal regions have different locations in the interphase nuclei of cold-blooded vertebrates. *Chromosoma* **2006**, *115*, 123–128. [[CrossRef](#)]

56. Dekker, J. GC- and AT-rich chromatin domains differ in conformation and histone modification status and are differentially modulated by Rpd3p. *Genome Biol.* **2007**, *8*, R116. [[CrossRef](#)]
57. Liu, J.; Ali, M.; Zhou, Q. Establishment and evolution of heterochromatin. *Ann. N. Y. Acad. Sci.* **2020**. [[CrossRef](#)]
58. Varriale, A.; Torelli, G.; Bernardi, G. Compositional properties and thermal adaptation of 18S rRNA in vertebrates. *RNA* **2008**, *14*, 1–9. [[CrossRef](#)]
59. Cruveiller, S.; Jabbari, K.; Clay, O.; Bernardi, G. Compositional gene landscapes in vertebrates. *Genome Res.* **2004**, *14*, 886–892. [[CrossRef](#)]
60. Di Filippo, M.; Bernardi, G. Mapping DNase I-hypersensitive sites on human isochores. *Gene* **2008**, *419*, 62–65. [[CrossRef](#)]
61. Di Filippo, M.; Bernardi, G. The early apoptotic DNA fragmentation targets a small number of specific open chromatin region. *PLoS ONE* **2009**, *4*, e5010. [[CrossRef](#)]
62. Guelen, L.; Pagie, L.; Brassat, E.; Meuleman, W.; Faza, M.B.; Talhout, W.; Eussen, B.H.; de Klein, A.; Wessels, L.; de Laat, W.; et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **2008**, *453*, 948–951. [[CrossRef](#)]
63. Meuleman, W.; Peric-Hupkes, D.; Kind, J.; Beaudry, J.B.; Pagie, L.; Kellis, M.; Reinders, M.; Wessels, L.; van Steensel, B. Constitutive nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res.* **2013**, *23*, 270–280. [[CrossRef](#)]
64. Bersaglieri, C.; Santoro, R. Genome Organization in and around the nucleolus. *Cells* **2019**, *8*, 579. [[CrossRef](#)]
65. Kumar, Y.; Sengupta, D.; Bickmore, W.A. Recent advances in the special organization of the mammalian genome. *J. Biosci* **2020**, *45*, 18. [[CrossRef](#)]
66. Dekker, J.; Rippe, K.; Dekker, M.; Kleckner, N. Capturing chromosome conformation. *Science* **2002**, *295*, 1306–1311. [[CrossRef](#)]
67. Gibcus, J.H.; Dekker, J. The hierarchy of 3D genome. *Mol. Cell.* **2013**, *49*, 773–782. [[CrossRef](#)]
68. Lieberman-Aiden, E.; van Berkum, N.L.; Williams, L.; Imakaev, M.; Ragozy, T.; Telling, A.; Amit, I.; Lajoie, B.R.; Sabo, P.J.; Dorschner, M.O.; et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **2009**, *326*, 289–293. [[CrossRef](#)]
69. Dixon, J.R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **2012**, *485*, 376–380. [[CrossRef](#)] [[PubMed](#)]
70. Nora, E.P.; Lajoie, B.R.; Schulz, E.G.; Giorgetti, L.; Okamoto, I.; Servant, N.; Piolot, T.; van Berkum, N.L.; Meisig, J.; Sedat, J.; et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*. **2012**, *485*, 381–385. [[CrossRef](#)]
71. Sexton, T.; Yaffe, E.; Kenigsberg, E.; Bantignies, F.; Leblanc, B.; Hoichman, M.; Parrinello, H.; Tanay, A.; Cavalli, G. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **2012**, *148*, 458–472. [[CrossRef](#)]
72. Rao, S.S.P.; Huntley, M.H.; Durand, N.C.; Stamenova, E.K.; Bochkov, I.D.; Robinson, J.T.; Sanborn, A.L.; Machol, I.; Omer, A.D.; Lander, E.S.; et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **2014**, *159*, 1665–1680. [[CrossRef](#)] [[PubMed](#)]
73. Liu, J.; Zhang, L.; Quan, H.; Tian, H.; Meng, L.; Yang, L.; Feng, H.; Gao, Y.Q. From 1D sequence to 3D chromatin dynamics and cellular functions: A phase separation perspective. *Nucl. Acids Res.* **2018**, *46*, 9367–9383. [[CrossRef](#)] [[PubMed](#)]
74. Quan, H.; Yang, Y.; Liu, S.; Tian, H.; Xue, Y.; Gao, Y.Q. Chromatin structure changes during various processes from a DNA sequence view. *Curr. Opin. Struct. Biol.* **2020**, *62*, 1–8. [[CrossRef](#)]
75. Tian, C.; Yang, Y.; Liu, S.; Quan, H.; Gao, Y.Q. Toward an understanding of the relation between gene regulation and 3D genome organization. *BioRxiv* **2020**. [[CrossRef](#)]
76. Aissani, B.; Bernardi, G. CpG islands, genes and isochores in the genomes of vertebrates. *Gene* **1991**, *106*, 185–195. [[CrossRef](#)]
77. Jabbari, K.; Bernardi, G. CpG doublets, CpG islands and Alu repeats in long human DNA sequences from different isochore families. *Gene* **1998**, *224*, 123–128. [[CrossRef](#)]
78. Zhang, L.; Dai, Z.; Yu, J.; Xiao, M. CpG-island-based annotation and analysis of human housekeeping genes. *Brief. Bioinform.* **2020**, *2020*, 515–525. [[CrossRef](#)]
79. Cammarano, R.; Costantini, M.; Bernardi, G. The isochore patterns of invertebrate genomes. *BMC Genom.* **2009**, *10*, 538–550. [[CrossRef](#)]
80. Wang, S.; Su, J.H.; Beliveau, B.J.; Bintu, B.; Moffitt, J.R.; Wu, C.T.; Zhuang, X. Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **2016**, *353*, 598–602. [[CrossRef](#)]
81. Kind, J.; Pagie, L.; de Vries, S.S.; Nahidiazar, L.; Dey, S.S.; Bienko, M.; Zhan, Y.; Lajoie, B.; de Graaf, C.A.; Amendola, M.; et al. Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell* **2015**, *163*, 134–147. [[CrossRef](#)]
82. Naumova, N.; Imakaev, M.; Fudenberg, G.; Zhan, Y.; Lajoie, B.R.; Mirny, L.A.; Dekker, J. Organization of the mitotic chromosome. *Science* **2013**, *342*, 948–953. [[CrossRef](#)]
83. Jabbari, K.; Chakraborty, M.; Wiehe, T. DNA sequence-dependent chromatin architecture and nuclear hubs formation. *Sci. Rep.* **2019**, *9*, 14646. [[CrossRef](#)]
84. Naughton, C.; Avlonitis, N.; Corless, S.; Prendergast, J.G.; Mati, I.K.; Eijk, P.P.; Cockroft, S.L.; Bradley, M.; Ylstra, B.; Gilbert, N. Transcription Forms and Remodels Supercoiling Domains Unfolding Large-Scale Chromatin Structures. *Nat. Struct. Mol. Biol.* **2013**, *20*, 387–395. [[CrossRef](#)] [[PubMed](#)]

85. Schwarzer, W.; Abdennur, N.; Goloborodko, A.; Pekowska, A.; Fudenberg, G.; Loe-Mie, Y.; Fonseca, N.A.; Huber, W.; Haering, C.H.; Mirny, L.; et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **2017**, *551*, 51–56. [[CrossRef](#)] [[PubMed](#)]
86. Rao, S.S.P.; Huang, S.C.; Glenn St Hilaire, B.; Engreitz, J.M.; Perez, E.M.; Kieffer-Kwon, K.R.; Sanborn, A.L.; Johnstone, S.E.; Bascom, G.D.; Bochkov, I.D.; et al. Cohesin loss eliminates all loop domains. *Cell* **2017**, *171*, 305–320. [[CrossRef](#)] [[PubMed](#)]
87. Nora, E.P.; Goloborodko, A.; Valton, A.L.; Gibcus, J.H.; Uebersohn, A.; Abdennur, N.; Dekker, J.; Mirny, L.A.; Bruneau, B.G. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **2017**, *169*, 930–944. [[CrossRef](#)]
88. Finn, E.H.; Pegoraro, G.; Brandão, H.B.; Valton, A.L.; Oomen, M.E.; Dekker, J.; Mirny, L.; Misteli, T. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **2019**, *176*, 1502–1515. [[CrossRef](#)]
89. Matthews, N.E.; White, R. Chromatin Architecture in the Fly: Living without CTCF/Cohesin Loop Extrusion? *BioEssays* **2019**, *41*, 1900048. [[CrossRef](#)]
90. Quinodoz, S.A.; Ollikainen, N.; Tabak, B.; Palla, A.; Schmidt, J.M.; Detmar, E.; Lai, M.M.; Shishkin, A.A.; Bhat, P.; Takei, Y.; et al. Higher order interchromosomal hubs shape 3D genome organization in the nucleus. *Cell* **2018**, *174*, 744–757. [[CrossRef](#)]
91. Banigan, E.J.; van den Berg, A.A.; Brandao, H.B.; Marko, J.F.; Mirny, L.A. Chromosome organization by one-sided and two-sided loop extrusion. *Elife* **2020**, *9*, e53558. [[CrossRef](#)]
92. Wang, Y.; Zhang, T.; Zhang, R.; van Schaik, T.; Zhang, L.; Sasaki, T.; Peric-Hupkes, D.; Chen, Y.; Gilbert, D.M.; van Steensel, B.; et al. SPIN reveals genome-wide landscape of nuclear compartmentalization. *BioRxiv* **2020**. [[CrossRef](#)]
93. Mourad, R. Studying 3D evolution using genomic sequences. *Bioinformatics* **2019**, *36*, 1367–1373. [[CrossRef](#)]
94. Hudson, A.P.; Cuny, G.; Cortadas, J.; Haschemeyer, A.E.V.; Bernardi, G. An analysis of fish genomes by density gradient centrifugation. *Eur. J. Biochem.* **1980**, *112*, 203–210. [[CrossRef](#)]
95. Costantini, M.; Bernardi, G. The short sequence design of isochores from the human genome. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 13971–13976. [[CrossRef](#)]
96. Arhondakis, S.; Auletta, F.; Bernardi, G. Isochores and the Regulation of Gene Expression in the Human Genome. *Genome Biol. Evol.* **2011**, *3*, 1080–1089. [[CrossRef](#)]
97. Satchwell, S.C.; Drew, H.R.; Travers, A.A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **1986**, *191*, 659–675. [[CrossRef](#)]
98. Widom, J. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **2001**, *34*, 269–324. [[CrossRef](#)]
99. Anderson, J.D.; Widom, J. Poly (dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol. Cell. Biol.* **2001**, *21*, 3830–3839. [[CrossRef](#)]
100. Sekinger, E.A.; Moqtaderi, Z.; Struhl, K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **2005**, *18*, 735–748. [[CrossRef](#)]
101. Segal, E.; Fondufe-Mittendorf, Y.; Chen, L.; Thåström, A.C.; Field, Y.; Moore, I.K.; Wang, J.P.Z.; Widom, J. A genomic code for nucleosome positioning. *Nature* **2006**, *442*, 772–778. [[CrossRef](#)]
102. Struhl, K.; Segal, E. Determinants of nucleosome positioning. *Nat. Struct. Mol. Biol.* **2013**, *20*, 267–273. [[CrossRef](#)]
103. Barbic, A.; Zimmer, D.P.; Crothers, D.M. Structural origins of adenine-tract bending. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 2369–2373. [[CrossRef](#)]
104. Fenouil, R.; Cauchy, P.; Koch, F.; Descostes, N.; Cabeza, J.Z.; Innocenti, C.; Ferrier, P.; Spicuglia, S.; Gut, M.; Gut, I.; et al. CpG islands and GC content dictate nucleosome depletion in a transcription-independent manner at mammalian promoters. *Genome Res.* **2012**, *22*, 2399–2408. [[CrossRef](#)]
105. Ricci, M.A.; Manzo, C.; García-Parajo, M.F.; Lakadamyali, M.; Cosma, M.P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **2015**, *160*, 1145–1158. [[CrossRef](#)]
106. Henikoff, S.; Ahmad, K. Nucleosomes remember where they were. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 20254–20256. [[CrossRef](#)]
107. D’Onofrio, G.; Ghosh, T.C.; Saccone, S. Different functional classes of genes are characterized by different compositional properties. *FEBS Lett.* **2007**, *581*, 5819–5824. [[CrossRef](#)]
108. Mirny, L.A.; Imakaev, M.; Abdennur, N. Two major mechanisms of chromosome organization. *Curr. Opin. Cell Biol.* **2019**, *58*, 142–152. [[CrossRef](#)]
109. Hon, C.C.; Carninci, P. ENCODE expanded. *Nature* **2020**, *583*, 685–686. [[CrossRef](#)]
110. Falk, M.; Feodorova, Y.; Naumova, N.; Imakaev, M.; Lajoie, B.R.; Leonhardt, H.; Joffe, B.; Dekker, J.; Fudenberg, G.; Solovei, I.; et al. Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **2019**, *570*, 395–399. [[CrossRef](#)]
111. Dolgin, E. A loop of faith. *Nature* **2017**, *544*, 284–286. [[CrossRef](#)]
112. Eyre-Walker, A.; Hurst, L.D. The evolution of isochores. *Nat. Rev. Genet.* **2001**, *2*, 549–555. [[CrossRef](#)]
113. Battulin, N.; Fishman, V.S.; Mazur, A.M.; Pomaznoy, M.; Khabarova, A.A.; Afonnikov, D.A.; Prokhortchouk, E.B.; Serov, O.L. Comparison of the three-dimensional organization of sperm and fibroblast genomes using the Hi-C approach. *Genome Biol.* **2015**, *16*, 77. [[CrossRef](#)]
114. Trifonov, E.N. Sequence-dependent deformational anisotropy of chromatin DNA. *Nucleic Acids Res.* **1980**, *8*, 4041–4053. [[CrossRef](#)] [[PubMed](#)]
115. Trifonov, E.N. The multiple codes of nucleotide sequences. *Bull. Mathem. Biol.* **1989**, *51*, 417–432. [[CrossRef](#)] [[PubMed](#)]

116. Todolli, S.; Perez, P.J.; Clauvelin, N.; Olson, W. Contributions of sequence to the higher-order structures of DNA. *Biophys. J.* **2017**, *112*, 416–426. [[CrossRef](#)] [[PubMed](#)]
117. Ramirez, F.; Bhardwaj, V.; Arrigoni, L.; Lam, K.C.; Grüning, B.A.; Villaveces, J.; Habermann, B.; Akhtar, A.; Manke, T. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat. Comm.* **2018**, *9*, 189. [[CrossRef](#)]
118. Gorkin, D.U.; Qiu, Y.; Hu, M.; Fletez-Brant, K.; Liu, T.; Schmitt, A.D.; Noor, A.; Chiou, J.; Gaulton, K.J.; Sebat, J.; et al. Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biol.* **2019**, *20*, 255. [[CrossRef](#)]
119. Fudenberg, G.; Kelley, D.R.; Pollard, K.S. Predicting 3D genome folding from DNA sequence. *Nat. Methods* **2020**. [[CrossRef](#)]
120. Schwessinger, R.; Gosden, M.; Downes, D.; Brown, R.; Telenius, J.; Teh, Y.W.; Lunter, G.; Hughes, J.R. DeepC: Predicting chromatin interactions using megabase scaled deep neural networks and transfer learning. *Nat. Methods* **2020**. [[CrossRef](#)]
121. Kasinathan, S.; Henikoff, S. Non-B-Form DNA is enriched at centromeres. *Mol. Biol. Evol.* **2018**, *35*, 949–962. [[CrossRef](#)]
122. Doolittle, W.F.; Sapienza, G. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **1980**, *284*, 601–603. [[CrossRef](#)]
123. Orgel, L.E.; Crick, F.H. Selfish DNA: The ultimate parasite. *Nature* **1980**, *284*, 604–607. [[CrossRef](#)]
124. Palazzo, F.; Gregory, R.T. The Case for Junk DNA. *PLoS Genet.* **2014**, *10*, e1004351. [[CrossRef](#)]
125. Graur, D. An upper limit on the functional fraction of the human genome. *Genome Biol. Evol.* **2017**, *9*, 1880–1885. [[CrossRef](#)]
126. Doolittle, W.F.; Brunet, D.P. On causal roles and selected effects: Our genome is mostly junk. *BMC Biol.* **2017**, *15*, 116. [[CrossRef](#)]
127. Kimura, M. Evolutionary rate at the molecular level. *Nature* **1968**, *217*, 624–626. [[CrossRef](#)]
128. Kimura, M. *The Neutral Theory of Molecular Evolution*; Cambridge University Press: Cambridge, UK, 1985.
129. Ohta, T. Slightly deleterious mutant substitutions in evolution. *Nature* **1973**, *246*, 96–98. [[CrossRef](#)]
130. Ohta, T. Near-neutrality in evolution of genes and gene regulation. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 16134–16137. [[CrossRef](#)]
131. Akdemir, K.; Victoria, T.L.; Chandran, S.; Li, Y.; Verhaak, G.R.; Beroukhi, R.; Campbell, P.J.; Chin, L.; Dixon, J.R.; Futreal, P.A.; et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat. Genet.* **2020**, *52*, 294–305. [[CrossRef](#)]