The formation of chromatin domains: a new model

Giorgio Bernardi

gbernardi@uniroma3.it

Science Department, Roma Tre University, Viale Marconi 446, 00146 Rome, Italy, and Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy.

Abstract

The mechanisms of formation of LADs, the lamina associated domains, and TADs, the topologically associating domains of mammalian chromatin, were investigated here by using as a starting point the observation that chromatin architecture relies on an isochore framework and by doing a new analysis of both isochore structure and the isochore/chromatin domain connection. This approach showed that LADs correspond to isochores from the very GC-poor, compositionally very homogeneous L1 family and from the "low-heterogeneity" L2 (or $L2^-$) sub-family; in fact, LADs are compositionally flat, flexible chromatin structures (because of the nucleosome depletion associated with the frequent oligo-A's) that attach themselves to the nuclear lamina in self-interacting clusters. In contrast, TADs correspond to the increasingly GC-richer isochores from the "high-heterogeneity" L2 (or $L2^+$) sub-family and from the H1, H2 and H3 families. These isochores, making the framework of the individual chromatin loops or of the chromatin loop ensembles of TADs, were found to consist of single or multiple GC peaks. The self-interacting single or multiple loops of TADs appear to be shaped by the property that accompany the increasing levels of GC and CpG islands in their isochore peak backbones, namely by an increasing bendability due to decreasing nucleosome density which is accompanied by decreasing supercoiling and increasing nuclease accessibility. In conclusion, chromatin architecture appears to be encoded and molded by isochores, the DNA units of genome organization. This "isochore encoding/molding model" of chromatin domains represents a paradigm shift compared to previously proposed models. Indeed, the latter only rely on the properties of architectural proteins, whereas the new model is essentially based on the physico-chemical properties of isochores and on their differential binding of nucleosomes.

Introduction

In interphase nuclei, chromatin comprises two sets of domains that are largely conserved in mammals: LADs, the lamina associated domains (~0.5Mb median size), that are scattered over all chromosomes and correspond to GC-poor sequences (1-3), and TADs, the topologically associating domains (0.2-2 Mb in size), a system of GC-rich loops (4-7); many TADs can be resolved into contact domains (0.185 Mb median size (6).

In spite of the recent, impressive advances in our understanding of chromatin domains (see refs. 8-14, for reviews), the problem of their formation mechanism(s) is still unsolved. Interesting models have been proposed (15-24), but no satisfactory solution has been reached so far. The currently predominant model for TAD formation is the "chromatin extrusion model" (18,19), which proposes that TADs emerge as a consequence of loop extrusion by loop extruding factors (including cohesin), which is limited by boundary elements (including CTCF).

Here, the problem of chromatin domain formation was approached by taking into account the observation that isochores (see ref. 25 for a review) make up the framework of chromatin domains (26) and by having a new look at isochore structure and at the isochore/chromatin domain connection. This approach was applied to human chromosome 21 which was chosen because 1) this chromosome is a good representative of human chromosomes, allowing an extension of the results to all other chromosomes (as investigated in ref. 26); and 2) is the smallest human chromosome, allowing a more expanded graphical presentation of data.

As far as nomenclature is concerned, although TADs comprise, by definition, all the topologically associating domains, in the context of this article TADs will indicate the chromatin domains other than LADs. The main reason for this choice is that the mechanisms of formation of the two sets of domains are different, even if based on the same fundamental DNA property.

Results

Isochore structure: a new analysis

Figure 1A shows the compositional profile of the DNA sequence from human chromosome 21 as seen through non-overlapping 100Kb windows. This window size was used because 100KB is a plateau value under which the composition of DNA segments show an increasing variance with decreasing size (27) due to several factors (for instance, the distribution of interspersed repeated sequences).

Figures 1B and 1C display the isochore profile as obtained from the chromosome 21 sequence using either a sliding window approach (28,29; Fig. 1B) or a fixed window approach (27, 30; Fig. 1C). Both approaches flatten the compositional profiles by averaging, in two different ways, the fluctuating values of the large regions characterized by "fairly homogeneous" composition, the isochores. In the case of the sliding window approach, remnants of the fluctuations can still be seen as small spikes in GC-rich regions, whereas in the case of the fixed window approach the fluctuations disappear because of the strict averaging procedure applied.

A new, simpler, in fact elementary, approach was used here, namely plotting the individual GC values of 100Kb DNA blocks as points. This approach was suggested by two recent results: 1) correlations hold between isochore properties and the properties of chromatin domains (31); and, more precisely, 2) the framework of TADs and LADs is made up by GC-rich and GC-poor isochores, respectively (26). One may therefore imagine a possible topological similarity between the flat structure of LADs and the loops of TADs on the one hand, and the compositionally flat GC-poor and the compositionally heterogeneous GC-rich isochores, respectively, on the other. If such is the case, clearly a simple 100Kb point-by-point profile of GC levels is preferable not only to both sliding and fixed window approaches, that flatten the compositional profile, but also to the color bar plot of 100Kb DNA segments (Fig. 1A) for graphical clarity reasons.

95 The point-by-point profile (see Fig. 1D) expectedly showed the compositionally flat
96 region 2 and the H1 and L2 peaks (**a** to **f**) of regions 1 and 3, that were already evident in Figs.
97 1A, 1B and 1C. It also led, however, to the discovery that the sequences of isochores from H1
98 (region 4), H2 (region 5) and H3 (region 6) isochores were not simply fluctuating within the
99 compositional borders of the corresponding families (see Supplementary Table S1), but
100 consisted, in fact, of sets of GC peaks. Upon close inspection, these very evident peaks may be
101 seen to correspond to the minute peaks of Fig. 1B, that were flattened by the sliding window
102 approach. Expectedly, the peaks of the point-by-point plots covered a broader GC range
103 compared with the flattened peaks of the sliding window approach, as shown by comparing Fig.
104 1D with Fig. 1B, and Fig. 1F with Fig. 1E (the high resolution compositional profiles of a multi-
105 peak H2 isochore from chromosome 20).
106 In purely compositional terms (see Fig. 2 for a larger-scale presentation of the data of Fig.
107 1D), three different situations were found: 1a) a series of single peaks (regions 1 and 3)
108 corresponding to an H1 isochore (**a**), and to several L2 isochores (**b** to **f**), in which latter case
109 very few points were slightly beyond the "fixed" isochore family borders, but still within the
110 "extended" borders of Supplementary Table S1; 1b) several very sharp H3 single peaks (region
111 6) that included sequences belonging to the H2 and even to the H1 family, in which case an
112 overall GC range of 18% was reached; 2) a very homogeneous L1 isochore (region 2), in which
113 the overall GC range barely reached 4% and all points were within the "fixed" GC borders of
114 isochore family L1; and 3) two series of GC-rich multi-peak isochores that belonged to H1
115 (region 4) and H2 (region 5) families in which, again, very few points were slightly beyond the
116 "fixed" isochore family borders. The striking difference between the compositional profile of L1
117 and H3 isochores is shown in Fig. 2A. Expectedly, when using a higher resolution windows
118 (50Kb; see Supplementary Figure S1) the compositional profiles of isochore peaks became

119 broader in the GC level gradients and more complex, because of the presence of interspersed
120 repeated sequences and CpG islands (see ref. 27).
121
122 *Isochores and LADs*
123 It is well established (see refs. 1-3) that LADs 1) may cover ~35% of the human genome;
124 2) comprise 1,100-1,400 discrete domains demarcated by CTCF sites and CpG islands; 3) have a
125 median size of ~0.5Mb; 4) are scattered over all chromosomes; 5) can be subdivided into cLADs,
126 *i.e.,* "constitutive" LADs present in the four cell types originally tested and fLADs "facultative"
127 LADs, only present in some cell types (in fact, only ~15% of the genome is involved in "stable
128 contacts" present in most cells); 6) are characterized, in the case of cLADs, by conserved
129 positions in syntenic regions of human and mouse chromosomes; 7) show a correspondence of
130 cLADs and ciLADs (the "constitutive inter-LADs") with GC-poor and GC-rich isochores,
131 respectively.
132 As shown in Figs. 3A (and 4A,4B), the major LAD of chromosome 21 corresponds to a
133 large L1 isochore (which, incidentally, includes an exceptional GC-poor interLAD; see also the
134 interval in the self-interacting domains corresponding to the L1 isochore in Fig. 4C). The other
135 LADs correspond to the L1 isochores that separate the L2 peaks (to be described below and in
136 the following section), and to a "valley" L2 isochore comprised between two H1 isochores (on
137 the right side of Fig. 3A). Moreover, two LADs flank the centromere; in fact, this appears to be
138 the rule for all human chromosomes, as judged by looking at the results of ref. 26.
139 In chromosome 20 (Fig. 3B), the largest LAD corresponds to an L2 isochore (interrupted
140 by an interLAD) while several other LADs correspond to L2 valley isochores flanked by H1
141 isochores; among faint LADs, one (extreme right) corresponds to an H1 isochore comprised
142 between two H2 isochores and two other ones flank the centromere. In the very GC-rich

143 chromosome 19 (Fig. 3C), two LADs correspond to two H1 isochores flanking an H2 isochore
144 and two other LADs correspond to L2 isochores flanking an H1 isochore; finally, two faint LADs
145 flank the centromere.

146 These results show that LADs correspond not only to L1 isochores that represent ~19% of
147 the genome (incidentally, not too far from the ~15% involved in "stable contacts"; see ref. 3), but
148 also to L2 isochores and even to H1 isochores in the rare case of very GC-rich chromosomes.

149 As far as L2 isochores are concerned, it appears (see Supplementary Table 1 and Fig. 1)
150 that 1) some isochores belong to a "low-heterogeneity" L2 sub-family that may be called $L2^-$,
151 show a flat profile (see, for example, the largest LAD of chromosome 20); and 2) some other
152 isochores belong to a "high-heterogeneity" L2, or $L2^+$, sub-family that are higher in average GC
153 and are in the shape of single peaks (see Figs. 1 A-D and 3A). Now, as shown in Fig. 3, $L2^-$
154 isochores correspond to LADs, whereas $L2^+$ isochores correspond to interLADs and TADs (see
155 the following section). The remaining L2 isochores are generally present as valleys flanked by
156 GC-richer isochores (see Fig. 3A,3B,3C; the relative amounts of L2 sub-families are presented in
157 Supplementary Table S1).

158

159 *Isochores and TADs*

160 It should be recalled, as a preliminary remark, that the isochores from the five families
161 (L1, L2, H1, H2 and H3) of the human genome (and other mammalian genomes; 32) are
162 characterized not only by increasing GC levels and different short-sequence frequencies, but also
163 by increasing compositional heterogeneities, increasing levels of CpG, CpG islands and Alu
164 sequences and by decreasing levels of LINE sequences and of 5mC/CpG ratios (27,33-37).
165 Moreover, at the chromatin level, GC increases are correlated with higher bendability (38),

166 higher nuclease accessibility (39,40), lower nucleosome density (41) and lower supercoiling
167 (42,43), all properties linked to DNA sequences.
168 The connection of the isochores of chromosome 21, as seen in Figs. 1D and 2, with
169 chromatin loops can be described as follows (see Fig. 4A,4B,4C): 1) regions 1 and 3 show a
170 series of H1 (a) and L2 (b to f) isochores in which latter case at least some of their single peaks
171 appear to correspond to individual self-interactions; 2) region 2 is the GC-poorest L1 isochore
172 which corresponds to two self-interactions (separated by an exceptional interLAD); 3) the multi-
173 peak H1 isochores of region 4 correspond to a large interLAD region and to several self-
174 interactions; the two short sequences X and Y, corresponding to LADs, separate region 4 from
175 regions 3 and 5; 4) the small multi-peak H2 isochore (region 5) seems to correspond to a single
176 self-interaction; this may be due to the dense packing of the peaks and/or to a lack of resolution;
177 5) a series of H3 isochores (red points comprised between two red arrows) correspond to a series
178 of self-interactions comprised between the two red lines on the heat map; in this case, the six H3
179 isochore peaks correspond to at least three chromatin loops. In conclusion, the two classes of
180 isochores, single-peak and multi-peak, essentially correspond to two classes, single-loop and
181 multi-loop, respectively, of TADs (both of which also show inter-chromosomal interactions; 26).
182 The correspondence between isochores peaks and self-interactions is improved at a higher
183 resolution of the heat map (compare the high-resolution Fig. S2A with the low resolution Fig.
184 S2B). Likewise, a very good match of isochore peaks with chromatin loops can be seen in the
185 high resolution heatmap of the multipeak H2 isochore of chromosome 20 (see Supplementary
186 Fig. S3).
187 A very interesting correlation is shown in Fig. 4D, in that regions 1 and 3 to 6 correspond
188 to A compartments (open chromatin) whereas region 2 and the short X and Y sequences
189 correspond to B compartments (closed chromatin; see ref. 15). More precisely, the A

190 compartment corresponds to multi-peak isochore TADs (regions 1,3,4,5,6), the B compartment to
191 individual LADs (region 2,X,Y), the former being more frequent in telomeric regions.
192
193 **Discussion and Conclusion**
194 *Encoding of chromatin domains by isochores*
195 Very recent investigations showed that GC-poor and GC-rich isochores should be
196 visualized as the framework of chromatin architecture or, in other words, as the DNA units that
197 underlie LADs and TADs, respectively (26). This was an important step towards the idea that
198 isochores encode chromatin domains. The present results provide a conclusive evidence for this
199 idea, by showing a precise match between the chromatin domains and the isochores of
200 chromosome 21 and by generalizing these results to all human chromosomes.
201 Indeed, the compositional profiles, the heatmaps and the LAD maps (26) show that: 1) the
202 isochores from the L1 family and the $L2^-$ sub-family correspond to LADs in all human
203 chromosomes; 2) $L2^+$ peaks emerging from an L1 background and corresponding to interLADs
204 and TADs are also found in other human chromosomes, although less frequently than in
205 chromosome 21; likewise, H3 peaks also corresponding to interLADs and TADs are present in
206 most human chromosomes; 3) the spikes of the compositional profiles of H1 and H2 isochores
207 of Fig. 1B, that reflect the peaks of Fig. 1D, are regularly present in H1 and H2 isochores from all
208 human chromosomes and correspond to the peaks of point-by-point profiles (Cozzi P. et al.,
209 paper in preparation) and to heat map interactions. This general match is important in that the
210 only alternative to the encoding proposed here is that the match of the thousands of LADs and
211 TADs with the corresponding isochores is just a coincidence (and this cannot be *quia absurdum*).
212
213 *Molding of chromatin domains by isochores*

214 The present results also solve an important open problem, namely the mechanism of
215 formation of chromatin domains. Indeed, LADs should be visualized as chromatin structures
216 corresponding to GC-poor isochores that are flexible, because of the local nucleosome depletions
217 linked to the richness of oligo-A sequences in the corresponding isochores (33,37,44,45; G.
218 Lamolle, H. Musto, G. Bernardi; paper in preparation). LADs only twist and bend in order to
219 adapt and attach themselves to (and even embed in) the lamina, which is reassembled after
220 mitosis (3). Expectedly, this leads to self-interactions (see Fig. 4), as well as to interactions with
221 other LADs from the same chromosomes (26; see, for example, the two LADs bracketed by
222 black lines in Fig. 4). In the case of TADs, the GC gradient within each GC-rich isochore peak is
223 accompanied by properties, increasing levels of CpG, CpG islands and Alu sequences, that lead
224 to increasing nucleosome depletion and bendability and decreasing supercoiling (38-43). These
225 factors constrain the corresponding chromatin to fold into loops.
226 The models for the formation of LADs and TADs developed in this investigation are
227 presented in Fig. 5, which stresses a keypoint, namely the central role played by the
228 compositional properties of isochores in the formation of chromatin domains. Indeed, the folding
229 model presented here clearly relies on isochore sequences, their nucleosome depletion and the
230 emerging local (in LADs) or extended (in TADs) flexibility of the chromatin fiber.
231 It should be stressed that this "isochore encoding/molding model" of chromatin domains
232 represents a paradigm shift compared to previously proposed models. Indeed, the latter only rely
233 on the properties of architectural proteins, whereas the new model is essentially based on the
234 physico-chemical properties of isochores and on their differential binding of nucleosomes.
235 The "isochore encoding/molding model" of chromatin domains is, however, compatible 1)
236 with both the requirements of CTCF binding to close chromatin loops into insulated TADs (46)
237 and the lack of such requirements (47); 2) with the interaction of topoisomerase II beta with

238 cohesin and CTCF at topological domain borders (48); 3) with an "insulation-attraction model"
239 of TAD formation (9) in which the insulation observed at TAD boundaries may result from
240 stiffness of the chromatin fiber caused by functional elements (CTCF binding sites, highly active
241 transcription starting sites etc) associated with increased nucleosome density and specific local
242 chromatin interactions due to "attractive forces" (not better specified but possibly linked to
243 supercoiling); 4) with the "chromatin extrusion model" (18,19), if the initial attachment of the
244 loop extruder (cohesin) were to coincide with the tips (highest GC levels) of TAD loops.
245
246 *The "isochore encoding/molding model" vs the "chromatin extrusion model"*
247 Although, as just mentioned, "the extrusion model" could overlap with the "isochore
248 endoding/molding model", a question may be raised about which one of the two models is better
249 supported by facts. An answer may come by considering what happens in the case of the "mitotic
250 memory", namely the rapid and precise re-establishment of the original interphase chromatin
251 domains at the exit from mitosis. In the first case, the basic information required for such quick
252 re-establishment is already present in the sequences of isochores (CTCF may also play an
253 important role in the process). In the second case, the formation of thousands of loops involves
254 the attachment of loop extruding factors (possibly at specific sequences as suggested above) and
255 an extrusion process, which requires a source of energy. It is obvious that the first model relying
256 on well known intrinsic physical properties of DNA is to be preferred to the second one which
257 involves thousands of conjectural loop extruders and an unknown source of energy.
258 A final point should be made to stress that the models under discussion here concern the
259 basic evolutionarily stable chromatin domains, since, as it is well known, epigenetic
260 modifications and environmental factors may cause changes in chromatin architecture; indeed,
261 while self-associating domains are stable in mammals, chromatin interactions within and between

262 domains may change during differentiation (49) and evolution (the latter subject will be
263 discussed elsewhere).
264
265 *Isochores as functional genome units*
266 The present results lead to a new vision of isochores since they not only correspond to a
267 fundamental level of genome structure and organization (50), but also to a set of functional
268 genome units that encode and mold LADs and TADs. Several observations support the above
269 conclusion, three of which are the following: 1) the evolutionary conservation of the isochore
270 patterns in mammals (32); 2) TADs from cells of adult organisms are basic units of replication
271 timing (51) and GC-rich and GC-poor isochores are replication units characterized by all early or
272 all late replicons (52); 3) alterations of the architecture of chromatin domains (both LADs and
273 TADs), known to lead to senescence and diseases (see the review papers cited in the
274 Introduction), are due to changes in their isochore framework. This was predicted by previous
275 investigations on "genomic diseases" (53,54), defined as diseases due to sequence alterations that
276 do not affect genes or classical regulatory sequences, but other sequences that "cause regional
277 changes in chromatin structure".
278 The fact that alterations in the chromatin architecture, not affecting genes or classical
279 regulatory sequences, lead to problems in transcription 1) represents the strongest and final
280 objection to the idea of "junk DNA" (see ref. 55, for a review); and 2) has practical implications:
281 indeed, screening even a small human population in terms of chromatin structure in view of
282 detecting "genomic diseases" is simply not feasible at least at the present time. The availability of
283 LADs and TADs maps along sequenced human chromosomes may allow, however, to link
284 alterations at the DNA sequence level with chromatin structure alterations. For instance, the maps

285 of insertions, deletions and SNPs of Venter's chromosomes (56) in combination with reference
286 chromatin structure maps, might lead to detect problems in Venter's chromatin domains.
287
288 *The large-scale organization of the human genome*
289 We can now consider a higher level of isochore and chromatin organization. At the DNA
290 level, two "genome spaces" were defined on the basis of gene density (57,58): the gene-poor
291 "genome desert" (L1+L2 isochores) and the gene-rich "genome core" (H1+H2+H3 isochores). In
292 the interphase nucleus, the chromatin corresponding to the genome core showed an internal
293 location and an open structure, whereas the chromatin corresponding to the genome desert
294 showed a peripheral location and a closed structure (see Table 1); moreover, the former showed a
295 preference for (generally GC rich) telomeric regions, the latter for centromeric regions, this
296 preference explaining the polarity of chromosomes in the nucleus (59,60).
297 The recently proposed chromosome compartments, A and B, characterized by open and
298 closed chromatin, respectively (15), show properties very similar to those just described. In
299 conclusion, the two compartments, A and B (see the compartment profile of chromosome 21 in
300 Fig. 4D) appear to correspond to the two genome spaces, the "genome core" and the "genome
301 desert" (see Table 1). This conclusion is supported by the comparison of sub-compartments with
302 isochore profiles (26).
303
304 *The genomic code*
305 The encoding of chromatin domains by isochores deserves the name of "genomic code".
306 This definition was originally coined (61,62; see also ref. 25) for two sets of compositional
307 correlations 1) those that hold among genome sequences (for instance, between coding and
308 contiguous non-coding sequences) and among the three codon positions of genes and that reflect

309 isochore properties; and 2) those that link isochores with all the structural/functional properties of
310 the genome (25,26,31), the latter now including the properties of TADs and LADS. Here it is
311 proposed that the definition of "genomic code" be applied to the encoding of chromatin domains
312 by isochores, since this is in fact the basis for the second set of the correlations just mentioned.
313 Interestingly, the genomic code may be visualized as the fourth, and last, pillar of molecular
314 biology, the first one being the double helix (1951-1953), the second the regulation of gene
315 expression in *E. coli* (1957-1961), and the third the genetic code (1961-1966). In contrast with
316 the other pillars, the genomic code took decades to be established.
317

318 **Acknowledgements**

319 The author thanks Paolo Ascenzi for hospitality, Giacomo Bernardi, Oliver Clay, and,
320 especially, Kamel Jabbari for critical reading, comments and discussions as well as Caterina
321 Nuvoli for excellent technical help. This research was supported by the Kimura Prize for
322 Molecular Evolution and Evolutionary Genomics conferred to the author (Tokyo, June 2016).
323
324

325 **References:**

326 1. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, et al. Domain organization of
327 human chromosomes revealed by mapping of nuclear lamina interactions. Nature. 2008;453:
328 948–51.
329 2. . Meuleman W, Peric-Hupkes D, Kind J, Beaudry JB, Pagie L, Kellis M, et al. Constitutive
330 nuclear lamina-genome interactions are highly conserved and associated with A/T-rich sequence.
331 Genome Res. 2013;23: 270–280.
332 3. Kind J, Pagie L, De Vries SS, Nahidiazar L, Dey SS, Bienko M, et al. Genome-wide Maps of
333 Nuclear Lamina Interactions in Single Human Cells. Cell. 2015;163: 134–147.
334 4. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian
335 genomes identified by analysis of chromatin interactions. Nature. 2012;485: 376–380.
336 5. 2. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Spatial partitioning
337 of the regulatory landscape of the X-inactivation centre. Nature. 2012;485: 381–385.
338 6. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map
339 of the human genome at kilobase resolution reveals principles of chromatin looping. Cell.
340 2014;159: 1665–1680.
341 7. Vietri Rudan M, Barrington C, Henderson S, Ernst C, Odom DT, Tanay A, et al. Comparative
342 Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture. Cell Rep.
343 2015;10: 1297–1309.
344 8. Bonev B, Cavalli G. Organization and function of the 3D genome. Nat Rev Genet. Nature
345 Research; 2016;17: 661–678.
346 9. Dixon JR, Gorkin DU, Ren B. Chromatin Domains: The Unit of Chromosome Organization. Mol
347 Cell. 2016;62: 668–680.
348 10. Ghirlando R, Felsenfeld G. CTCF: Making the right connections. Genes and Development. 2016.
349 pp. 881–891.
350 11. Gonzalez-Sandoval A, Gasser SM. On TADs and LADs: Spatial Control Over Gene Expression.
351 Trends Genet. 2016;
352 12. Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene
353 Regulation. Annu Rev Genomics Hum Genet. 2016;17: annurev-genom-083115-022339.

354 13. Pueschel R, Coraggio F, Meister P, Ahmed K, Dehghani H, Rugg-Gunn P, et al. From single
355 genes to entire genomes: the search for a function of nuclear organization. Development. Oxford
356 University Press for The Company of Biologists Limited; 2016;143: 910–23.
357 14. Solovei I, Thanisch K, Feodorova Y. How to rule the nucleus: divide et impera. Curr Opin Cell
358 Biol. 2016;40: 47–59.
359 15. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al.
360 Comprehensive mapping of long-range interactions reveals folding principles of the human
361 genome. Science. 2009;326: 289–93.
362 16. Alipour E, Marko JF. Self-organization of domain structures by DNA-loop-extruding enzymes.
363 Nucleic Acids Res. Oxford University Press; 2012;40: 11202–12.
364 17. Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, et al. Complexity of chromatin
365 folding is captured by the strings and binders switch model. Proc Natl Acad Sci U S A. 2012;109:
366 16173–16178.
367 18. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of
368 Chromosomal Domains by Loop Extrusion. Cell Rep. 2016;15: 2038–2049.
369 19. Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, et al. Chromatin
370 extrusion explains key features of loop and domain formation in wild-type and engineered
371 genomes. Proc Natl Acad Sci USA. National Acad Sciences; 2015;112: E6456–E6465.
372 20. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. Cell.
373 2016;164: 1110–1121.
374 21. Di Pierro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JN. Transferable model for
375 chromosome architecture. Proc Natl Acad Sci. 2016; 201613607.
376 22. Rowley MJ, Corces VG. The three-dimensional genome: Principles and roles of long-distance
377 interactions. Current Opinion in Cell Biology. 2016. pp. 8–14.
378 23. Ulianov S V, Khrameeva EE, Gavrilov AA, Flyamer IM, Kos P, Mikhaleva EA, et al. Active
379 chromatin and transcription play a key role in chromosome partitioning into topologically
380 associating domains. Genome Res. Cold Spring Harbor Laboratory Press; 2016;26: 70–84.
381 24. Vietri Rudan M, Hadjur S. Genetic Tailors: CTCF and Cohesin Shape the Genome During
382 Evolution. Trends Genet. 2015;31: 651–660.

383 25. Bernardi G (2004, reprinted in 2005) *Structural and evolutionary genomics: natural selection in*
384 *genome evolution.* (Elsevier, Amsterdam). This book is freely available at
385 www.giorgiobernardi.eu
386 26. Jabbari K, Bernardi G. An isochore framework underlies chromatin architecture. Plos One
387 http://dx.doi.org/10.1371/journal.pone.0168023
388 27. Costantini M, Clay O, Auletta F, Bernardi G. An isochore map of human chromosomes. Genome
389 Res. 2006;16: 536–541.
390 28. Pavlícek A, Paces J, Clay O, Bernardi G. A compact view of isochores in the draft human
391 genome sequence. FEBS Lett. 2002;511: 165–9.
392 29. Pačes J, Zíka R, Pačes V, Pavlíček A, Clay O, Bernardi G. Representing GC variation along
393 eukaryotic chromosomes. Gene. 2004;333: 135–141.
394 30. Cozzi P, Milanesi L, Bernardi G. Segmenting the human genome into isochores. Evol
395 Bioinforma. 2015;11: 253–261.
396 31. Bernardi G. Chromosome architecture and genome organization. PLoS One. 2015;10.
397 doi:10.1371/journal.pone.0143739
398 32. Costantini M, Cammarano R, Bernardi G. The evolution of isochore patterns in vertebrate
399 genomes. BMC Genomics. 2009;10: 146. doi:10.1186/1471-2164-10-146
400 33. Costantini M, Bernardi G. The short-sequence designs of isochores from the human genome.
401 Proc Natl Acad Sci U S A. 2008;105: 13971–6.
402 34. Meunier-Rotival M, Soriano P, Cuny G, Strauss F, Bernardi G. Sequence organization and
403 genomic distribution of the major family of interspersed repeats of mouse DNA. Proc Natl Acad
404 Sci U S A. 1982;79: 355–9.
405 35. Soriano P, Meunier-Rotival M, Bernardi G. The distribution of interspersed repeats is
406 nonuniform and conserved in the mouse and human genomes. Proc Natl Acad Sci U S A.
407 1983;80: 1816–20.
408 36. Varriale A, Bernardi G. Distribution of DNA methylation, CpGs, and CpG islands in human
409 isochores. Genomics. 2010;95: 25–28.
410 37. Arhondakis S, Auletta F, Bernardi G. Isochores and the regulation of gene expression in the
411 human genome. Genome Biol Evol. 2011;3: 1080–9.

412 38. Vinogradov AE. DNA helix: The importance of being GC-rich. Nucleic Acids Research. 2003.
413 pp. 1838–1844.
414 39. Di Filippo M, Bernardi G. Mapping DNase-I hypersensitive sites on human isochores. Gene.
415 2008;419:62-65.
416 40. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible
417 chromatin landscape of the human genome. Nature. 2012;489: 75–82.
418 41. Fenouil R, Cauchy P, Koch F, Descostes N, Cabeza JZ, Innocenti C, et al. CpG islands and GC
419 content dictate nucleosome depletion in a transcription-independent manner at mammalian
420 promoters. Genome Res. 2012;22: 2399–2408.
421 42. Naughton C, Avlonitis N, Corless S, Prendergast JG, Mati IK, Eijk PP, et al. Transcription forms
422 and remodels supercoiling domains unfolding large-scale chromatin structures. Nat Struct Mol
423 Biol. 2013;20: 387–95.
424 43. Benedetti F, Dorier J, Burnier Y, Stasiak A. Models that include supercoiling of topological
425 domains reproduce several known features of interphase chromosomes. Nucleic Acids Res.
426 2014;42: 2848–55.
427 44. Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, et al. The DNA-
428 encoded nucleosome organization of a eukaryotic genome. Nature. 2009;458: 362–366.
429 45. Frenkel ZM, Bettecken T, Trifonov EN. Nucleosome DNA sequence structure of isochores.
430 BMC Genomics. 2011;12: 203.
431 46. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, et al. Target degradation of
432 CTCF decouples local insulation of chromosome domains from higher-order genome
433 compartmentalization. doi: https://doi.org/10.1101/095802
434 47. Kubo N, Ishii H, Gorkin D, Meitinger F, Xiong X, Fang R, et al. Preservation of Chromatin
435 Organization after Acute Loss of CTCF in Mouse Embryonic Stem Cells. bioRxiv. Cold Spring
436 Harbor Labs Journals; 2017; 118737.
437 48. Uuskülа-Reimand L, Hou H, Samavarchi-Tehrani P, Rudan MV, Liang M, et al. Topoisomerase
438 II beta interacts with cohesin and CTCF at topological domain borders. Genome Biol. 2016; 17:
439 1–22.
440 49. Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, et al. Chromatin
441 architecture reorganization during stem cell differentiation. Nature. 2015;518: 331–336.

50. Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, et al. The mosaic genome of warm-blooded vertebrates. Science. American Association for the Advancement of Science; 1985;228: 953–8.
51. Rivera-Mulia JC, Gilbert DM. Replicating Large Genomes: Divide and Conquer. Mol Cell. 2016;62: 756–765.
52. Costantini M, Bernardi G. Replication timing, chromosomal bands, and isochores. Proc Natl Acad Sci U S A. 2008;105: 3433–3437.
53. Bernardi G. The neoselectionist theory of genome evolution. Proc Natl Acad Sci U S A. 2007;104: 8385–90.
54. Al Mahmud A, Amore G, Bernardi G. Compositional genome contexts affect gene expression control in sea urchin embryo. PLoS One. 2008;3. doi:10.1371/journal.pone.0004025
55. Palazzo AF, Gregory TR. The Case for Junk DNA. PLoS Genet. 2014;10. doi:10.1371/journal.pgen.1004351
56. Costantini M, Bernardi G. Mapping insertions, deletions and SNPs on Venter's chromosomes. PLoS One. 2009;4. doi:10.1371/journal.pone.0005972
57. Bernardi G. The Vertebrate Genome : Isochores and Evolution. Mol Biol Evol. 1993;10: 186–204.
58. Bernardi G. Isochores and the evolutionary genomics of vertebrates. Gene. 2000. 3–17.
59. Sadoni N, Langer S, Fauth C, Bernardi G, Cremer T, Turner BM, et al. Nuclear organization of mammalian genomes: Polar chromosome territories build up functionally distinct higher order compartments. J Cell Biol. 1999;146: 1211–1226.
60. Saccone S, Federico C, Bernardi G. Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. Gene. 2002;300: 169–78.
61. Bernardi G. Le génome des vertébrés : organisation, fonction et évolution. Biofutur. 1990. 94 : 43-46.
62. Bernardi G, Bernardi G. Compositional properties of nuclear genes from cold-blooded vertebrates. J. Mol. Evol. 1986. 33:57-67.
63. Naumova N, Imakaev M, Fudenberg G, Zhan Y, Lajoie BR, Mirny LA, et al. Organization of the mitotic chromosome. Science. 2013;342: 948–53.

471 64. Consortium IHGS. Initial sequencing and analysis of the human genome. Nature. 2001;409: 860–
472 921.
473 65. Bernardi G. Misunderstandings about isochores. Part 1. Gene. 2001. pp. 3–13.
474 66. Jabbari K, Nürnberg P. A genomic view on epilepsy and autism candidate genes. Genomics.
475 2015. 108:31-36.
476
477

478 **Figure Legends**

479 **Figure 1. A:** Compositional profile (from ref. 30) of human chromosome 21 (release hg38) as seen
480 through non-overlapping 100-Kb windows. DNA stretches from isochore families L1 to H3 are
481 represented in different colors, deep blue, light blue, yellow, orange, red, respectively. The left-side
482 ordinate values are the minima GC values between the isochore families listed on the right side
483 (see Supplementary Table S1); **a-f** correspond to an H1 peak (**a**) and to several $L2^+$ peaks (**b** to **f**;
484 the thin yellow bars in peaks **c**, **e** and **f** correspond to single 100Kb blocks that are assigned to the
485 H1 family if using "fixed" isochore borders (as shown in the Figure), but to L2 isochores if
486 "extended" borders (see Supplementary Table 1) are used. Six regions numbered 1 to 6 (top and
487 bottom of the Figure) of the compositional profile are separated by vertical lines or double lines, X
488 and Y.

489 **B**. Isochore profile of human chromosome 21 using the matched b37 assembly of ref. 6 and a
490 sliding window approach (28,29) with "fixed" isochore borders (from ref. 26). The color
491 convention is as in Fig. 1A. This profile is slightly more extended on the centromeric (left) side
492 than that of Figs. 1A and 1C.

493 **C.** Isochore profile of human chromosome 21 (release hg38) using a non-overlapping 100Kb
494 window and the isoPlotter program (from ref. 30).

495 **D**. GC levels of 100Kb windows of human chromosome 21. This figure shows that individual
496 isochores from the $L2^+$ to H3 families are in the form of peaks. The GC gradients of the peaks do
497 not appear in a clear way in the standard presentation of compositional profiles of chromosomes
498 (Fig. 1A), except for the broad, isolated H1 (**a**) and $L2^+$ peaks (**b** to **f**). Upon close inspection,
499 however, the peaks of the H1, H2 and H3 isochores of this Figure show a correspondence with the
500 small peaks of isochores belonging to H1, H2 and H3 family present in Fig. 1B. Blue and red

501 arrows delimit regions 2 and 6, respectively. Black, blue and red arrows, as well as X,Y double
502 lines (see Fig. 4 and its legend for additional information) separate regions 3,4 and 5; horizontal
503 lines correspond to minima GC values between isochore families (see legend of Fig. 1A and
504 Supplementary Table 1).

505 **E.F.** The compositional map of a 2.1Mb region of human chromosome 20, using a sliding window
506 approach (from ref. 28) and the "extended" family borders of Supplementary Table 1 (Fig. 1E), is
507 aligned with the corresponding "point-by-point map" (Fig. 1F, in the Fig. 1D style). The
508 correspondence of five peaks and valleys from the latter (red lines) precisely correspond to the
509 peaks and valleys of the compositional profile; several additional peaks and valleys show a less
510 precise, yet unequivocal correspondence (blue lines). The GC range of the peaks from the bottom
511 profile, 1F, is much larger (about double) than the corresponding one of the top profile, 1E. The
512 whole region may be visualized as a multipeak H2 isochore on the basis of the bottom profile (see
513 also Supplementary Fig. S3).

514

515 **Figure 2A.** GC levels of 100Kb windows from L1 and H3 regions of human chromosome 21
516 comprised between the blue and red arrows (regions 2 and 6) of Fig. 1D, respectively, are
517 displayed at a higher magnification. Horizontal lines correspond to "fixed" minima GC values
518 between isochore families (see Supplementary Table 1). The abscissa scale is in 100Kb units.

519 **B.C**. GC levels of 100Kb windows from the regions 1,3,4 and 5 of human chromosome 21. See the
520 legend of Fig. 2A for other indications. **C.** Region 4 is delimited by two GC-poor sequences, X and
521 Y, that correspond to LADs (see Fig. 4 and its legend for additional information).

522

523 **Figure 3. A.** The isochore profile of human chromosome 21 is compared with the (inverted) LAD
524 profile (from ref. 3) to show the correspondence of LADs 1) with L1 isochores (blue lines; two
525 broken blue lines bracket the largest L1 isochore, in which case the LAD is (exceptionally)
526 interrupted by an interLAD); 2) with one L2 "valley" isochore (blue line, last on the right side);
527 and 3) with two LADs flanking the centromere. Four "high heterogeneity" $L2^+$ "peak" isochores
528 (red lines) correspond to interLADs, or TADs; the H1 peak right of the centromere corresponds to
529 the same interLAD as the first $L2^+$ peak. Two LADs flank the centromere. The multicolored bar on
530 the right is the color code for isochore families.

531 **B.** The isochore profile of human chromosome 20 (see also legend of Fig. 3A) is compared with
532 the (inverted) LAD profile (from ref. 3) to show the correspondence of LADs with a large low-
533 heterogeneity L2 ($L2^-$) isochore (bracketed by broken blue lines; on the left of the panel) which
534 includes an interLAD (red line) and with several L2 "valley" isochores (blue lines), as well as the
535 correspondence of interLADs with GC-rich isochores (red lines); two faint LADs flank the
536 centromere.

537 **C.** The isochore profile of human chromosome 19 (see also legend of Fig. 3A) is compared with
538 the (inverted) LAD profile (from ref. 3) to show that two LADs correspond to H1 isochores (blue
539 lines) flanking a H2 isochore, an interLAD (red line); two other LADs (on the left) correspond to
540 L2 "valley" isochores (blue lines); two faint LADs flank the centromere.

541 **Figure 4.** The compositional profile of human chromosome 21 (**B,** from Fig. 1D), is compared 1)
542 with the corresponding LAD profile (**A**; an inverted scale is shown on the top right of the figure);
543 2) with the heat map of chromatin interactions (**C**; from ref. 26) and with Fig. 1A on the right side;
544 3) with the A/B compartment profile (**D**; data for mid G1 from ref. 63). Regions 1 and 3
545 correspond to multiple interactions and to open chromatin; region 2 to multiple interactions and

546 closed chromatin, as well as to an interLAD. Two double vertical black lines, X and Y,
547 corresponding to LADs (characterized by both self- and intra-chromosomal interactions) and to
548 closed chromatin separate region 4 a multi-peak H1 isochore block corresponding to multiple
549 interactions and open chromatin, from regions 3 and 5. The small multi-peak H2 isochores (region
550 5) corresponds to a single self-interaction (see text) and to open chromatin (see Text). A telomeric
551 block of H3 isochores (region 6) defined by red points, arrows and lines corresponds to several (at
552 least three) self-interactions on the heat map and to open chromatin.
553

554 **Figure 5**. Models for the formation of LADs and TADs. Three chromatin fibers are taken into
555 consideration: **A**. a GC-poor chromatin fiber corresponding to an L1 isochore; blue bar) bounded
556 by CTCF binding sites (green) attaches to the lamina, forming a LAD; the wavy profile indicates
557 the physical adaptation by bending and twisting to (and embedding in) the lamina due to
558 nucleosome depletion associated with oligo-A sequences (yellow stripes), as well as the self-
559 interactions. **B**. a GC-rich chromatin fiber corresponding to an H3 isochore folds upon itself
560 forming a single-loop TAD bounded by CTCF-binding sites (green); yellow to red color indicate
561 an increasing GC level which is responsible for the folding due to nucleosome depletion associated
562 with increasing GC, CpG islands and Alu sequences. **C**. A GC-rich chromatin fiber corresponding
563 to an H1 (or H2) isochore characterized by two GC peaks folds to form a TAD which comprises
564 two loops and two contact domains.

565

566

TABLE 1. GENOME ORGANIZATION: DNA AND CHROMATIN (a)		
2 Genome spaces (57,58)	**Genome core** **GC-rich, gene-rich**	**Genome desert** **GC-poor, gene-poor**
2 Nuclear locations (59,60)	**Internal**	**Peripheral**
2 Chromatin structure (60)	**Open**	**Closed**
2 Replication timings (51,52)	**Early**	**Late**
2 Genome compartments (15)	**A, open chromatin**	**B, closed chromatin**
5 Isochore families (25)	**L2^{+}, H1, H2, H3**	**L1, L2^{-}**
2 Chromatin domains (3-7)	**TADs**	**LADs**

567
568 (a) see heading "The large-scale organization of the Human genome" in the Discussion section
569 of the manuscript for a description of this Table.

570

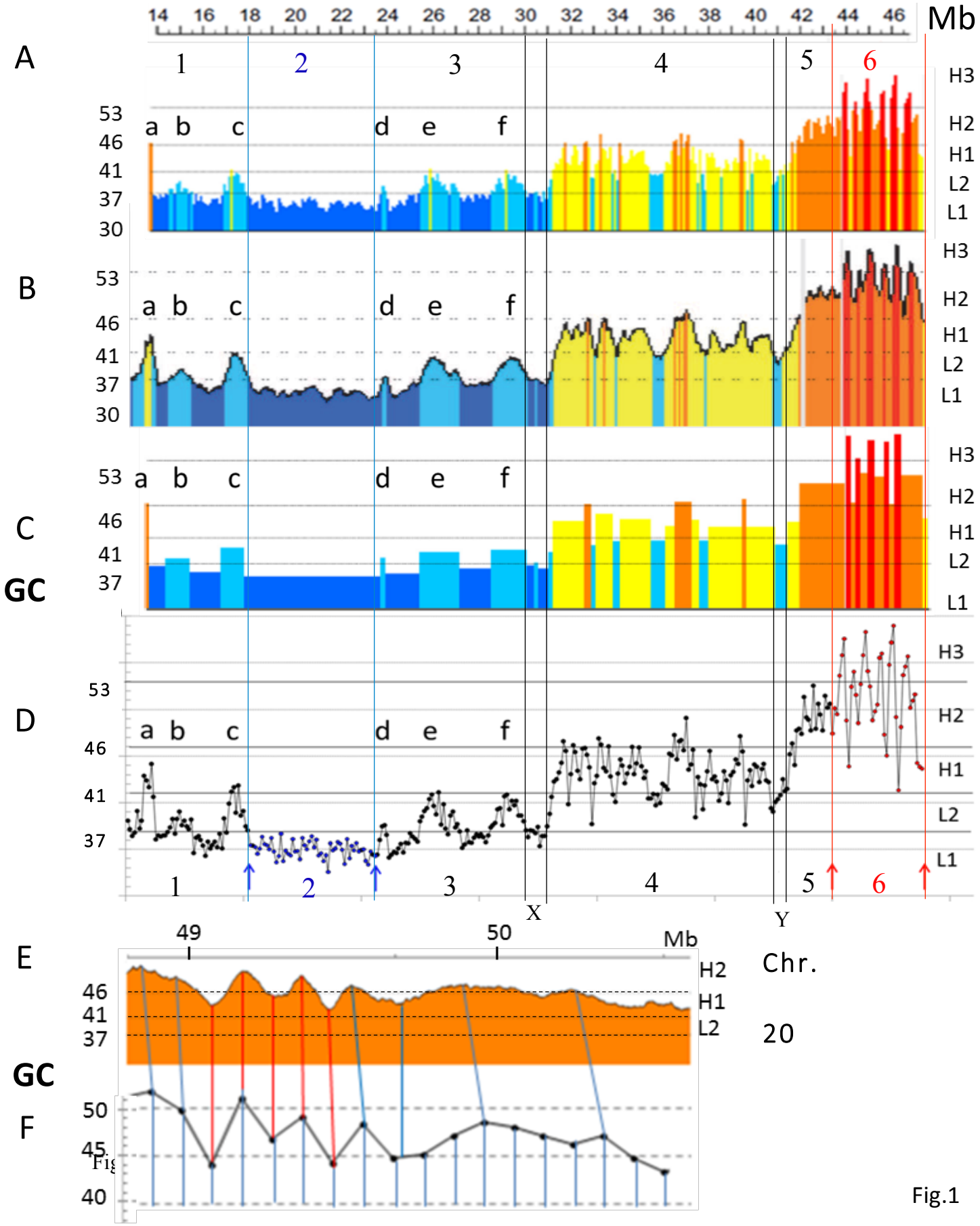
Mb
14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46
A
B
C
GC
D
E
F
1 2 3 4 5 6
a b c d e f
53 46 41 37 30
H3 H2 H1 L2 L1
X Y
49 50
Chr.
20
50 45 40

Fig.1

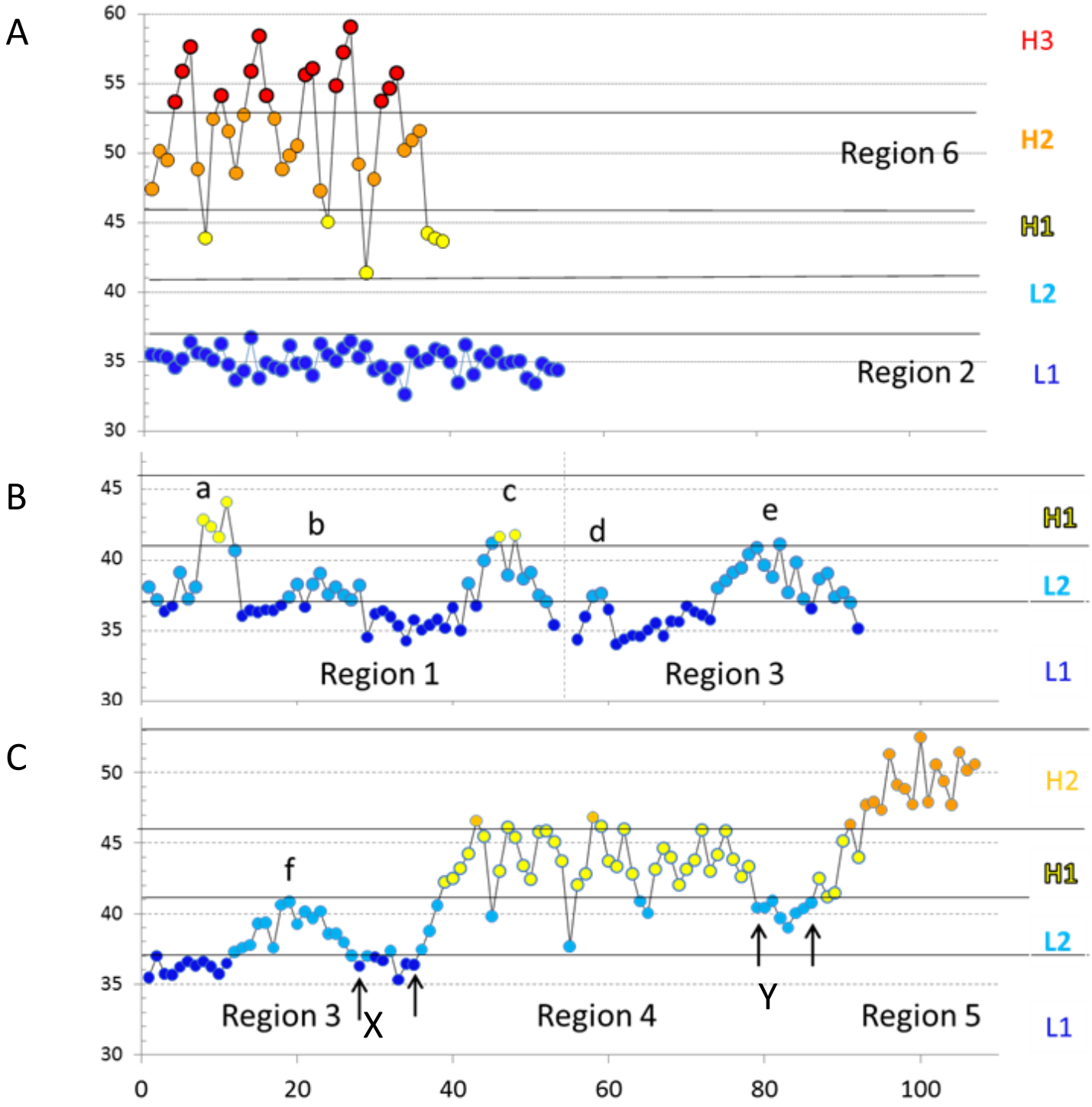
A
B
C
H3
H2
H1
L2
L1
Region 6
Region 2
a
b
c
d
e
Region 1
Region 3
f
X
Y
Region 3
Region 4
Region 5
60
55
50
45
40
35
30
0
20
40
60
80
100

Fig. 2

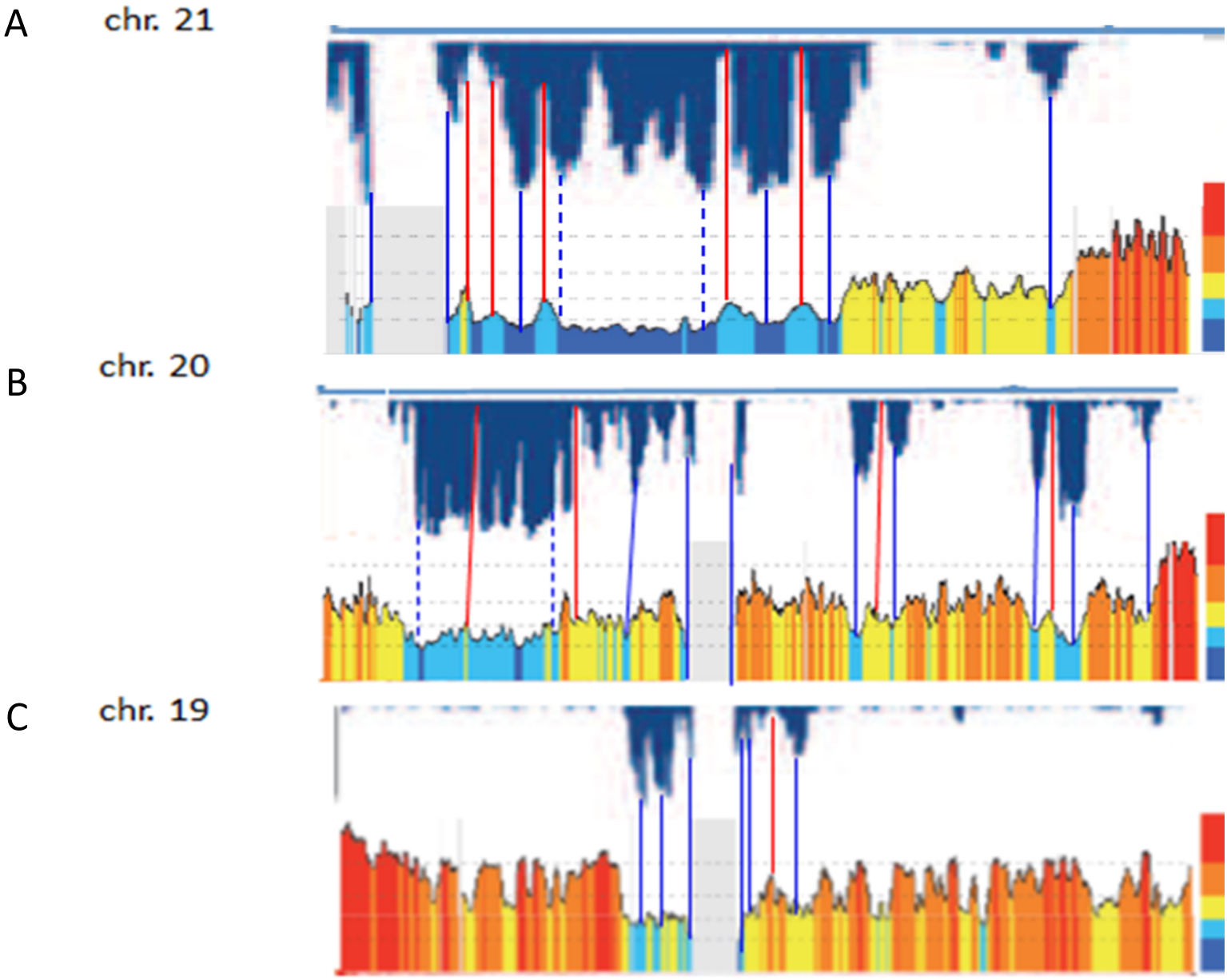

Fig. 3

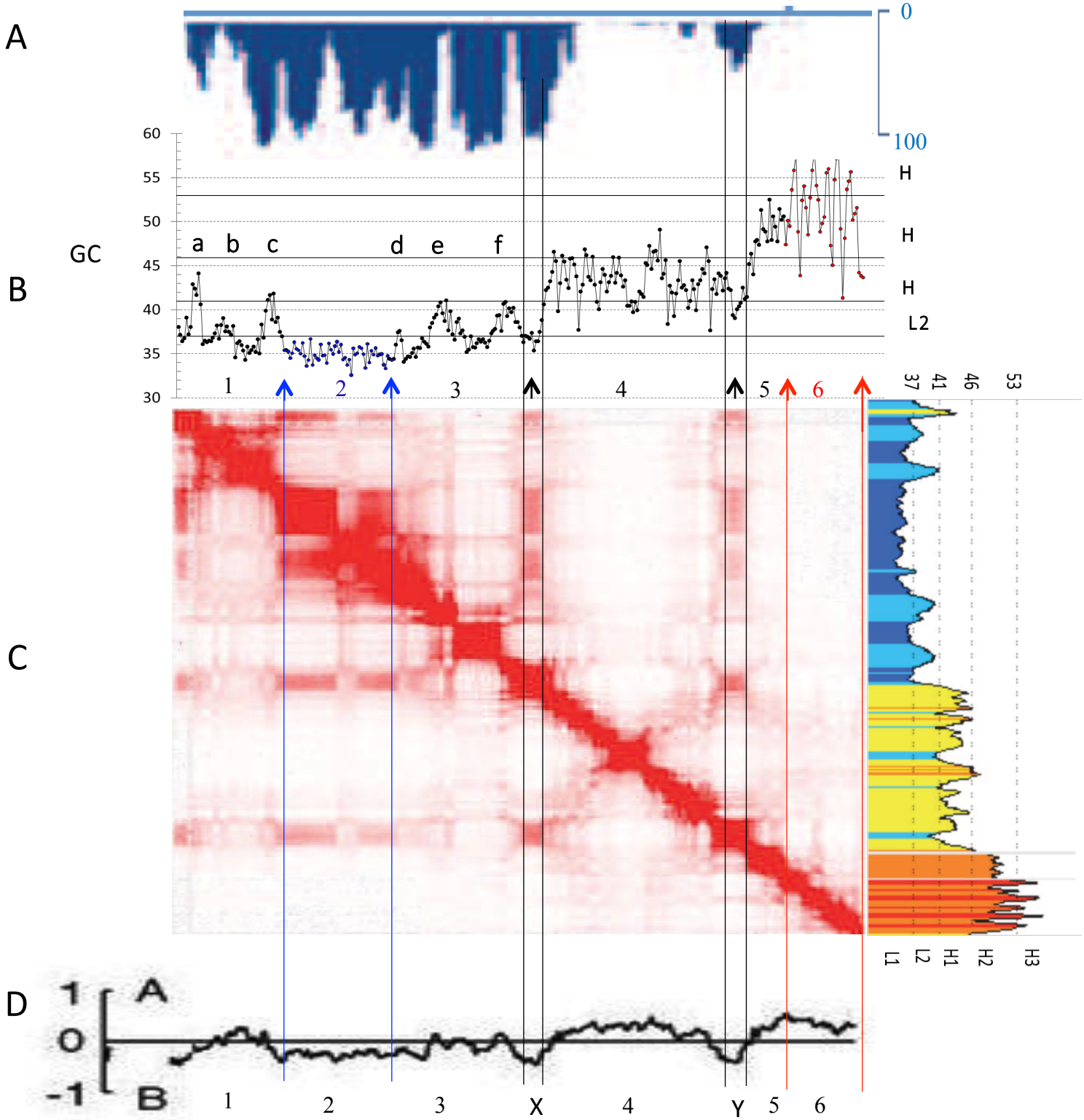
A
0
100
B
GC
60
55
50
45
40
35
30
a b c d e f
H
H
H
L2
1 2 3 4 5 6
37 41 46 53
C
L1 L2 H1 H2 H3
D
1
0
-1
A
B
1 2 3 X 4 Y 5 6

Fig. 4

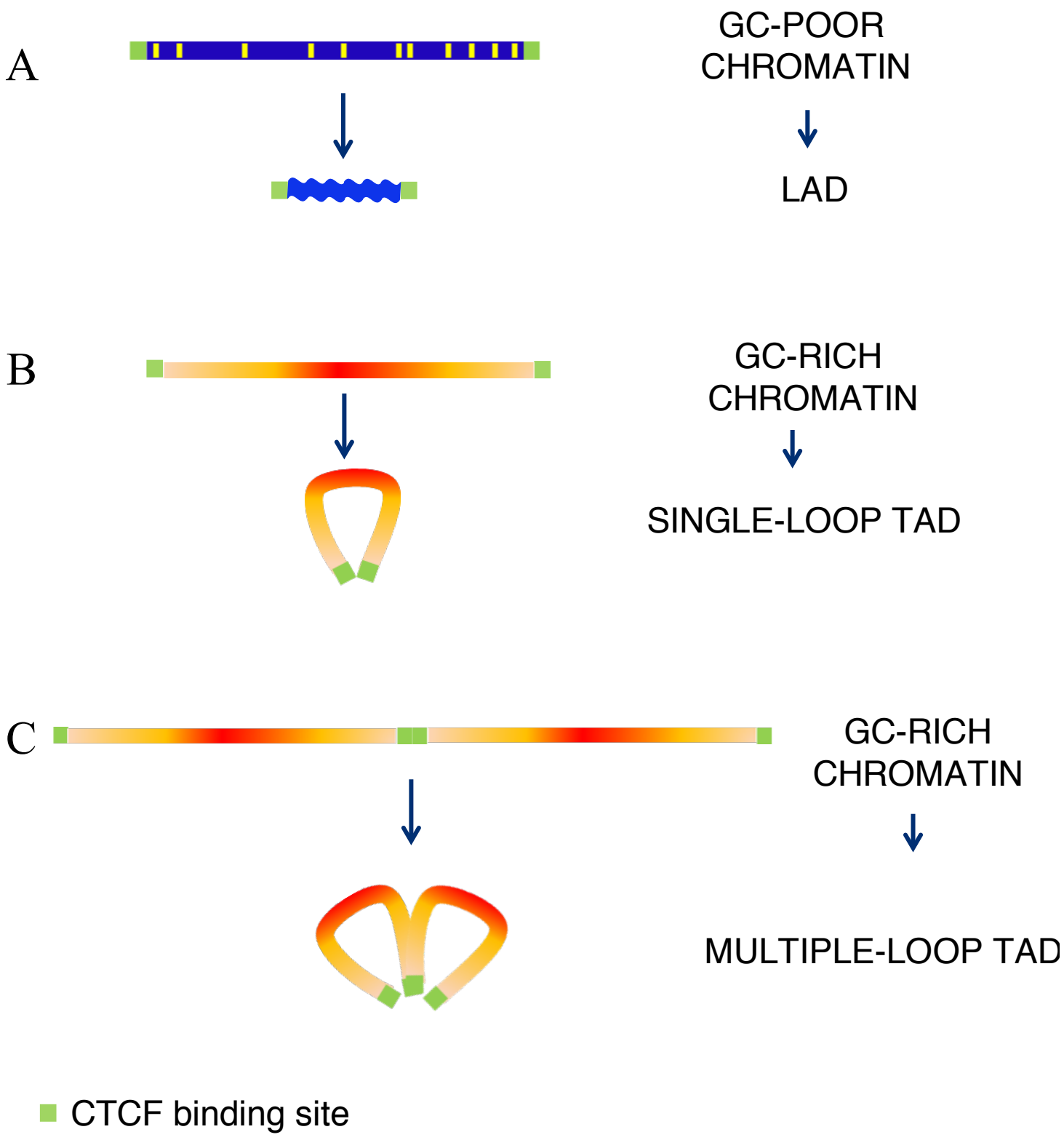
A
GC-POOR
CHROMATIN
LAD
B
GC-RICH
CHROMATIN
SINGLE-LOOP TAD
C
GC-RICH
CHROMATIN
MULTIPLE-LOOP TAD
CTCF binding site

Fig. 5